

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 21, 2004, 02:05:42 ; Search time 11154 Seconds
(without alignments)
11391.904 Million cell updates/sec

Title: US-09-900-751-1
Perfect score: 3106
Sequence: 1 catgtagacggtgccgg.....ttaaaaaaaaaaaaaaa 3106

Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 2888711 seqs, 2045481386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl.*

1: gb.ba.*

2: gb.htg.*

3: gb.in.*

4: gb.om.*

5: gb.ov.*

6: gb.pat.*

7: gb.ph.*

8: gb.pl.*

9: gb.pr.*

10: gb.ro.*

11: gb.sts.*

12: gb.sy.*

13: gb.un.*

14: gb.vi.*

15: em.ba.*

16: em.fun.*

17: em.hum.*

18: em.in.*

19: em.mu.*

20: em.om.*

21: em.or.*

22: em.ov.*

23: em.pat.*

24: em.ph.*

25: em.pl.*

26: em.ro.*

27: em.sts.*

28: em.un.*

29: em.vi.*

30: em.htg.hum.*

31: em.htg.inv.*

32: em.htg.other.*

33: em.htg.mus.*

34: em.htg.pln.*

35: em.htg.rod.*

36: em.htg.mam.*

37: em.htg.vrt.*

38: em.sy.*

39: em.htgo.hum.*

40: em.htgo.mus.*

41: em.htgo.other.*

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	3106	100.0	3106	6	AX395268	Sequence
2	3106	100.0	3106	10	AF042822	Mus muscu
3	3102.8	99.9	3248	10	BC005496	Mus muscu
4	2595	83.5	3174	10	AB049189	Rattus no
5	2313.6	74.5	2568	10	AB037898	Rattus no
6	1905.6	61.4	3273	9	BC030532	Homo sapi
7	1889.8	60.8	3149	9	AF118224	Homo sapi
8	1884.8	60.7	3120	9	AF133086	Homo sapi
9	1883.2	60.6	3128	9	AB030036	Homo sapi
10	1883.2	60.6	3147	6	AR081724	Sequence
11	1883.2	60.6	3147	6	AR229704	Sequence
12	1883.2	60.6	3147	6	AR229712	Sequence
13	1883.2	60.6	3147	6	AR229797	Sequence
14	1883.2	60.6	3147	6	AX207945	Sequence
15	1883.2	60.6	3147	9	AF057145	Homo sapi
16	1530.8	49.3	2900	6	AR081725	Sequence
17	1530.8	49.3	2900	6	AR229705	Sequence
18	1530.8	49.3	2900	9	HSU20428	Human
19	1175.8	37.9	2148	9	BC005826	Homo sapi
20	945.4	30.4	1823	9	BC018146	Homo sapi
21	809.6	26.1	3487	5	AB038498	Xenopus l
22	701.2	22.6	1553	6	AR263832	Sequence
23	623.2	20.1	225649	2	AC114542	Mus muscu
24	556.4	17.9	726	6	AX473052	Sequence
25	551.8	17.8	723	6	E13204	Human CDNA
26	359.6	11.6	236387	2	AC096042	Rattus no
27	359.6	11.6	261906	2	AC112632	Rattus no
28	359.6	11.6	277797	2	AC128347	Rattus no
29	306.8	9.9	434	6	AR273737	Sequence
30	306.8	9.9	434	6	AR277318	Sequence
31	306.8	9.9	434	6	AX368770	Sequence
32	265	8.5	591	6	AX261852	Sequence
33	244.8	7.9	3183	10	BC029645	Mus muscu
34	242.4	7.8	3194	9	AY055384	Homo sapi
35	240.8	7.8	2449	9	HSA319876	Homo sapi
36	240.8	7.8	2672	6	AX207905	Sequence
37	240.8	7.8	3104	6	AX207903	Sequence
38	239.2	7.7	1327	6	AX538208	Sequence
39	239.2	7.7	2409	6	AX360098	Sequence
40	239.2	7.7	3143	6	AX538206	Sequence
41	238	7.7	240079	2	AC128364	Rattus no
42	212.2	6.8	174526	9	AP001183	Homo sapi
43	212.2	6.8	182736	2	AC021672	Homo sapi
44	212.2	6.8	190314	9	AC019227	Homo sapi
45	206	6.6	734	6	AX375702	Sequence

ALIGNMENTS

RESULT 1	AX395268	AX395268	3106 bp	DNA	linear	PAT 18-MAY-2002
LOCUS	Sequence 5 from Patent WO0203787.					
DEFINITION	AX395268					
ACCESSION	AX395268					
VERSION	AX395268.1	GI:21066293				
KEYWORDS	Mus musculus (house mouse)					
SOURCE	Mus musculus					
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.					
REFERENCE	Allen, K.D. and Leviten, M.W.					
AUTHORS	Transgenic mice containing targeted gene disruptions					
TITLE	Patent: WO 0203787-A 5 17-JAN-2002;					
JOURNAL						

KRASGVQELKLRILTHPSNDFDYDIALLELEKSEVYSTVVRICLPDATHVFF
 AGKAIWVGTHGKEGTGALLQKGEIRVINQTCEDLMPEQITPRMWCVGLSGGV
 DCSQSDSGFPLSSAEKDGKRFVSWGEGCAQRNKGVTYTRLPVVRDVIKEHTGY"
 692 a 862 c 897 g 655 t
 BASE COUNT
 ORIGIN

```

Query Match      100.0%; Score 3106; DB 10; Length 3106;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 3106; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY	1	CATGTTAGACGGTGTCCCGGAGGGACACAGCGTCTGTAGAACGGCGATTCGAGACCGCCAAAA	60
Db	1		
QY	61	CATGTTAGACGGTGTCCCGGAGGGACACAGCGTCTGTAGAACGGCGATTCGAGACCGCCAAAA	60
Db	61		
QY	61	CNATGGGTAGCAATCGGGGCGCAAGGCGCGAGGGGGCTCTCAGGACTTCGGCGGGGAC	120
Db	61		
QY	61	CNATGGGTAGCAATCGGGGCGCAAGGCGCGAGGGGGCTCTCAGGACTTCGGCGGGGAC	120
Db	61		
QY	121	TCAAGTACAACCTCCCGGCTTAGAGAACATGAATGGCTTTGAGGAGGGTGTGGAGTTCCTGC	180
Db	121		
QY	121	TCAAGTACAACCTCCCGGCTTAGAGAACATGAATGGCTTTGAGGAGGGTGTGGAGTTCCTGC	180
Db	121		
QY	181	CTCGGAACAATCCGAAGAAAGTGGAGAGAGGAGGCGCCAGCGCTGGGTGGTGTGGTGG	240
Db	181		
QY	181	CTCGGAACAATCCGAAGAAAGTGGAGAGAGGAGGCGCCAGCGCTGGGTGGTGTGGTGG	240
Db	181		
QY	241	CAGTGTCTTCAGCTTCCTCTTGCTCTCCCTCATGGCTGGCTGGCTGTGGTGTGGCATCTCC	300
Db	241		
QY	241	CAGTGTCTTCAGCTTCCTCTTGCTCTCCCTCATGGCTGGCTGGCTGTGGTGTGGCATCTCC	300
Db	241		
QY	301	ATTATCGGAATGTGGGGTTCAAAAGTCTTTCAATGGCCATCTCAGGATCACAATAGAGA	360
Db	301		
QY	301	ATTATCGGAATGTGGGGTTCAAAAGTCTTTCAATGGCCATCTCAGGATCACAATAGAGA	360
Db	301		
QY	361	TCTTTCTGATCGGTATGAGAACTCCACTCCAAGAGTTTATCAGCTGGCAGCGAGG	420
Db	361		
QY	361	TCTTTCTGATCGGTATGAGAACTCCACTCCAAGAGTTTATCAGCTGGCAGCGAGG	420
Db	361		
QY	421	TGAAGGAGGCGCTGAAGCTGTGTACAATGAAGTCCCTGTCTCTGGGTCCCTACCAAGA	480
Db	421		
QY	421	TGAAGGAGGCGCTGAAGCTGTGTACAATGAAGTCCCTGTCTCTGGGTCCCTACCAAGA	480
Db	421		
QY	481	AGTTCGGCTGTAACTGCCCTTCAGTAGGGGCACTGTCTATCGGCTACTACTGTGTGAGTTCA	540
Db	481		
QY	481	AGTTCGGCTGTAACTGCCCTTCAGTAGGGGCACTGTCTATCGGCTACTACTGTGTGAGTTCA	540
Db	481		
QY	541	GNATCCCCCAACCTGGCAGAGAGGTTGATTCGGCCATGGCTGTGGAGCGAGTTGTAA	600
Db	541		
QY	541	GNATCCCCCAACCTGGCAGAGAGGTTGATTCGGCCATGGCTGTGGAGCGAGTTGTAA	600
Db	541		
QY	601	CATTGCCACCCGAGCAGCGGCACCTGAATCCTTCGTGCTAAACATCTGTGGTGGCGCTCC	660
Db	601		
QY	601	CATTGCCACCCGAGCAGCGGCACCTGAATCCTTCGTGCTAAACATCTGTGGTGGCGCTCC	660
Db	601		
QY	661	CCATTGACCCCAAGATGTGTGAGAGGACTCAGGACAAACAGCTGCAGTTTGGCCTCGATG	720
Db	661		
QY	661	CCATTGACCCCAAGATGTGTGAGAGGACTCAGGACAAACAGCTGCAGTTTGGCCTCGATG	720
Db	661		
QY	721	CCCATGGTGCACAGTGTACAAGCTTCACTACCCCTGACTCCCAACAGTGCCTACCCGG	780
Db	721		
QY	721	CCCATGGTGCACAGTGTACAAGCTTCACTACCCCTGACTCCCAACAGTGCCTACCCGG	780
Db	721		
QY	781	CGCATCCCCGCTGCCAGTGGGTCTCGGGGGGAGCGCGACTGTGTCTGAGGCTCACTC	840
Db	781		
QY	781	CGCATCCCCGCTGCCAGTGGGTCTCGGGGGGAGCGCGACTGTGTCTGAGGCTCACTC	840
Db	781		
QY	841	TCCGAAGCTTTGATGTTCGCTCCCTGTGTAGAGCATGGCAGTGCCTGATG	900
Db	841		
QY	841	TCCGAAGCTTTGATGTTCGCTCCCTGTGTAGAGCATGGCAGTGCCTGATG	900
Db	841		
QY	901	ATAGCTGTAGCCCATGTGAAACCCACAGCTGTGGTGGGCTGTGTGGCACTTCTCAACCT	960
Db	901		
QY	901	ATAGCTGTAGCCCATGTGAAACCCACAGCTGTGGTGGGCTGTGTGGCACTTCTCAACCT	960
Db	901		

Qy	961	CCTACAACTGAC	CTTTCCTCTCTCTCCAGAAAGTCTTCTCTGTGACGGTCAATAACCAATA	1020
Db	961	CCTACAACTGAC	TTTTCCTCTCTCTCCAGAAAGTCTTCTCTGTGACGGTCAATAACCAATA	1020
Qy	1021	CTGACGGCGACAT	CTCTGGCTTTGAGGCCACTTTCTTCAGCTGCCCAAGATGAGCAGCT	1080
Db	1021	CTGACGGCGACAT	CTCTGGCTTTGAGGCCACTTTCTTCAGCTGCCCAAGATGAGCAGCT	1080
Qy	1081	GTGGCGCTTTT	TGAGTGACACCAAGGGGACATTTAGCAGCCCTTACTATCCAGGGCCACT	1140
Db	1081	GTGGCGCTTTT	TGAGTGACACCAAGGGGACATTTAGCAGCCCTTACTATCCAGGGCCACT	1140
Qy	1141	ACCGGCCACATCA	CTACTGCACATGGAATACAAGTGGCCCAACACCGGAAAGTGAAG	1200
Db	1141	ACCGGCCACATCA	CTACTGCACATGGAATACAAGTGGCCCAACACCGGAAAGTGAAG	1200
Qy	1201	TGGCGTTCAAAC	CTTCTATCTGGTGGACCCCAACGTACCAAGTGGGCTCTCTGCACCAAGG	1260
Db	1201	TGGCGTTCAAAC	CTTCTATCTGGTGGACCCCAACGTACCAAGTGGGCTCTCTGCACCAAGG	1260
Qy	1261	ACTATGTGGAGAT	CAACGGGGAGAGTACTGCGGTGAGAGGTCCCAGTTTGTGTGAGCA	1320
Db	1261	ACTATGTGGAGAT	CAACGGGGAGAGTACTGCGGTGAGAGGTCCCAGTTTGTGTGAGCA	1320
Qy	1321	GCAACGACGCA	AGATTACAGTCCACTTTCATTTCTGATCACTCGTACACGGACACCGGCT	1380
Db	1321	GCAACGACGCA	AGATTACAGTCCACTTTCATTTCTGATCACTCGTACACGGACACCGGCT	1380
Qy	1381	TCTTAGCTGAGT	ACCTCTCTTACGATCCAAAGACCCGGTCCCGAGGATTTTCATGTGCA	1440
Db	1381	TCTTAGCTGAGT	ACCTCTCTTACGATCCAAAGACCCGGTCCCGAGGATTTTCATGTGCA	1440
Qy	1441	AGACTGACGGT	GCATCCGAAAGAACTCGCTGCGACGGCTGGGAGAGTCTGCCCGGAT	1500
Db	1441	AGACTGACGGT	GCATCCGAAAGAACTCGCTGCGACGGCTGGGAGAGTCTGCCCGGAT	1500
Qy	1501	ATAGTGATGAG	CGTTTACTGCGGATGCAATGCCACCCACAGTTTCAGTGCACCAAAACAGT	1560
Db	1501	ATAGTGATGAG	CGTTTACTGCGGATGCAATGCCACCCACAGTTTCAGTGCACCAAAACAGT	1560
Qy	1561	TCGTGCAAGCC	CTCTTCTGGTCTGTGACNGTGTCAAAGTGTGGGGACGGAGTGACG	1620
Db	1561	TCGTGCAAGCC	CTCTTCTGGTCTGTGACNGTGTCAAAGTGTGGGGACGGAGTGACG	1620
Qy	1621	AGGAGGCTGC	AGCTGTCTGTGGGATTTCAAGTGTTCCAATGGGAAGTGTCTCCCTC	1680
Db	1621	AGGAGGCTGC	AGCTGTCTGTGGGATTTCAAGTGTTCCAATGGGAAGTGTCTCCCTC	1680
Qy	1681	AGAGCCAGAAG	TGTAATGGGAAGGACAACTGTGGAGATGGGTCTGACAGGCTTCATGTG	1740
Db	1681	AGAGCCAGAAG	TGTAATGGGAAGGACAACTGTGGAGATGGGTCTGACAGGCTTCATGTG	1740
Qy	1741	ACAGCTGGAAT	GTCTTCTTGACCAATAATACCTACCGCTGCCCAATATGGCTCTGTG	1800
Db	1741	ACAGCTGGAAT	GTCTTCTTGACCAATAATACCTACCGCTGCCCAATATGGCTCTGTG	1800
Qy	1801	TGAGCAAGGCA	ACCTCTGATGTGATGGGAAGCAGCATGTAGCATGGCTCCGATGAGA	1860
Db	1801	TGAGCAAGGCA	ACCTCTGATGTGATGGGAAGCAGCATGTAGCATGGCTCCGATGAGA	1860
Qy	1861	AAACTGTGACT	GTGGGCTGCGATCTTTACCAAAAGGCTCGCGTGGTGGTGGACCA	1920
Db	1861	AAACTGTGACT	GTGGGCTGCGATCTTTACCAAAAGGCTCGCGTGGTGGTGGACCA	1920
Qy	1921	ATCGGACGAGG	CGAGTGGCCCTGGCAGTGAGCCTCCACGCCCTGGGCCAGGGCCACT	1980
Db	1921	ATCGGACGAGG	CGAGTGGCCCTGGCAGTGAGCCTCCACGCCCTGGGCCAGGGCCACT	1980
Qy	1981	TGTGTGGGCT	CGCTCACTCTCTCTGATGGGTGGTCTCTGAGCTCATATGCTTTCAGG	2040
Db	1981	TGTGTGGGCT	CGCTCACTCTCTCTGATGGGTGGTCTCTGAGCTCATATGCTTTCAGG	2040

2041 ATGACAAAATTTCAAGTACTCAGACTACAGATGTGACGGCTTCTCTGGTGTCTGG 2100
2041 ATGACAAAATTTCAAGTACTCAGACTACAGATGTGACGGCTTCTCTGGTGTCTGG 2100
2101 ACCAGACAGCGCAGTCCCTCTGGGTGAGAGCTGAAGCTCAACAGTATCATCCCC 2160
2101 ACCAGACAGCGCAGTCCCTCTGGGTGAGAGCTGAAGCTCAACAGTATCATCCCC 2160
2161 ACCCTTCTTCAATGATTTACCTTTCGACTATGACATCGCTTCTGGAGCTGAGAGT 2220
2161 ACCCTTCTTCAATGATTTACCTTTCGACTATGACATCGCTTCTGGAGCTGAGAGT 2220
2221 CGTGGAGTACAGACCGCTGTGCGCCCATCTCCCTGCTGATGCTACCATCTCTTCC 2280
2221 CGTGGAGTACAGACCGCTGTGCGCCCATCTCCCTGCTGATGCTACCATCTCTTCC 2280
2281 CTGCTGCAAGGCCATCTGGTTCACAGCTGGGGGCACACAAAAGAGGGAGGTACCGGAG 2340
2281 CTGCTGCAAGGCCATCTGGTTCACAGCTGGGGGCACACAAAAGAGGGAGGTACCGGAG 2340
2341 CGCTGATCTCTGCAAGGGTGAGATCGGTGTCTATCAACACAGACCATCTGTGAGGACCTCA 2400
2341 CGCTGATCTCTGCAAGGGTGAGATCGGTGTCTATCAACACAGACCATCTGTGAGGACCTCA 2400
2401 TGCCGACAGCATCACCCAGCAATGATGTGTGGGTTCCTCAGTGGGGGTGTGACT 2460
2401 TGCCGACAGCATCACCCAGCAATGATGTGTGGGTTCCTCAGTGGGGGTGTGACT 2460
2461 CTTGCCAGGTGACTCTGTGGTGGCCCTTGTCAAGCGGGGAAAGATGGCGATGTTC 2520
2461 CTTGCCAGGTGACTCTGTGGTGGCCCTTGTCAAGCGGGGAAAGATGGCGATGTTC 2520
2521 AGCTGTGTGTGAGTGGGGTGAAGGCTGCGCTCAGAGGAAACAAGCCAGGCGTGTACA 2580
2521 AGCTGTGTGTGAGTGGGGTGAAGGCTGCGCTCAGAGGAAACAAGCCAGGCGTGTACA 2580
2581 CAAGGCTCTCTGTAGTTTGGGACTGGATCAAGAGCACATGGGGTATAGCAGCATGGAC 2640
2581 CAAGGCTCTCTGTAGTTTGGGACTGGATCAAGAGCACATGGGGTATAGCAGCATGGAC 2640
2641 AGACAGCGGACCAACACACCCAGAGGATGCGGACATGCACACCTGGATACAGGAGAG 2700
2641 AGACAGCGGACCAACACACCCAGAGGATGCGGACATGCACACCTGGATACAGGAGAG 2700
2701 GAACACTGACGACATTTATGCTGTGGCTCCCTCCCTCCCTCCCTCCCTCCCTGTAAC 2760
2701 GAACACTGACGACATTTATGCTGTGGCTCCCTCCCTCCCTCCCTCCCTGTAAC 2760
2761 GCATCCTTAGACTCAGAGTTCTTCAAGTGGGACCCCTCAAGATTTGAGAGAGAAC 2820
2761 GCATCCTTAGACTCAGAGTTCTTCAAGTGGGACCCCTCAAGATTTGAGAGAGAAC 2820
2821 TTGCGTCTAGCGGCCAGCTGTGGGGCAAGGGTTTGTATGCGAGCTTCCCTCTAGCC 2880
2821 TTGCGTCTAGCGGCCAGCTGTGGGGCAAGGGTTTGTATGCGAGCTTCCCTCTAGCC 2880
2881 CTGAGCTGGTGAAGATGATGTGTCCGAGAGCTGCTTCCAACTGCTATTGAGCTCCC 2940
2881 CTGAGCTGGTGAAGATGATGTGTCCGAGAGCTGCTTCCAACTGCTATTGAGCTCCC 2940
2941 GGGAGCCCTATGGAGAGGGGCTCAGGTGCTCTTTTCAAGNAGCGCCAGCCCTAGGA 3000
2941 GGGAGCCCTATGGAGAGGGGCTCAGGTGCTCTTTTCAAGNAGCGCCAGCCCTAGGA 3000
3001 ACCCGAAGAGAGTGGTACTAAGGCTGAATTTTGTCTGTGCGAGGGTGGGTAT 3060
3001 ACCCGAAGAGAGTGGTACTAAGGCTGAATTTTGTCTGTGCGAGGGTGGGTAT 3060
3061 TTGAGAGTAAACATTTATTTCTTTTAAAAAATAAAAAA 3106
3061 TTGAGAGTAAACATTTATTTCTTTTAAAAAATAAAAAA 3106

RESULT 3
BC005496
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

BC005496 3248 bp mRNA linear ROD 16-APR-2003
Mus musculus suppression of tumorigenicity 14 (colon carcinoma),
mRNA (CDNA clone MGC:7395 IMAGE:3488059), complete cds.
BC005496.1 GI:13529565
MGC.

Mus musculus (house mouse)

Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 3248)

Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D.,
Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K.,
Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F.,
Diachenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L.,
Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L.,
Scheetz, T.E., Brownstein, M.J., Udwin, T.B., Toshiyuki, S.,
Carninci, P., Prange, C., Raha, S.S., Loquelliano, N.A., Peters, G.J.,
Abramson, R.D., Mullahy, S.J., Bosak, S.A., McEwan, P.J.,
McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S.,
Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W.,
Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A.,
Fahy, J., Helton, E., Kettman, M., Madan, A., Rodriguez, S.,
Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y.,
Bouffard, M.C., Blakesley, R.W., Touchman, J.W., Green, E.D.,
Dickson, G.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M.,
Butterfield, Y.S., Krzywinski, M.I., Skalska, U., Smalish, D.E.,
Schnerch, A., Schein, J.E., Jones, S.J. and Marra, M.A.
Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)

22388257

12477932

2 (bases 1 to 3248)

Strausberg, R.

Direct Submission

Submitted (27-MAR-2001) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA

NIH-MGC Project URL: <http://mgc.nci.nih.gov>

Contact: MGC help desk

Email: cgabs@mail.nih.gov

Tissue Procurement: Lothar Hennighausen Ph.D., Robin Humphreys

cDNA Library Preparation: Life Technologies, Inc.

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILNL)

DNA Sequencing by: Sequencing Group at the Stanford Human Genome

Center, Stanford University School of Medicine, Stanford, CA 94305

Web site: <http://www-shgc.stanford.edu>

Contact: (Dickson, Mark) mdpaxil.stanford.edu

Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers, R. M.

Clone distribution: MGC clone distribution information can be found

through the I.M.A.G.E. Consortium/ILNL at: <http://image.llnl.gov>

Series: IRAC Plate; 7 Row: i Column: 16

This clone was selected for full length sequencing because it

passed the following selection criteria: Hexamer frequency ORF

analysis.

Location/Qualifiers

1. 3248

/organism="Mus musculus"

/mol_type="mRNA"

/strain="C57BL/6J"

/db_xref="taxon:10090"

/clone="MGC:7395 IMAGE:3488059"

/tissue_type="Mammary tumor. WAP-TGF alpha model. 7 months

old, gross tissue."

/clone_lib="NCI CGAP_Mam5"

/lab_host="DH10B"

FEATURES
source

gene	/note="Vector: pCMV-SPORT6"		Db	651	GCATCCCCCACACCTGGCAGAGAGGTTGATCGCGCCATCGGCTGTGGAGCGAGTTGTAA		710			
	1. .3248									
CDS		/gene="St14"	QY	601	CATTGCCACCCCGAGCAGCGGCACTGAAATCTCTTGCTGCTAAACATCTGTGGTGGCCTTCC		660			
		/note="synonym: Epithin"								
		/db_xref="LocusID:19143"			Db	711		CATTGCCACCCCGAGCAGCGGCACTGAAATCTCTTGCTGCTAAACATCTGTGGTGGCCTTCC		770
		/db_xref="MG1:133881"								
	173_ .2740		QY	661	CCATTGACCCCAAGATGCTGCAGAGGACTCAGACAAACAGCTGCAGTTTGGCCCTGCATG		720			
	/codon_start=1									
	/product="suppression of tumorigenicity 14 (colon carcinoma)"		Db	771	CCATTGACCCCAAGATGCTGCAGAGGACTCAGGACAAACAGCTGCAGTTTGGCCCTGCATG		830			
	/protein_id="AAH05496.1"		QY	721	CCCATGGTGCAGCAGTGACACGCTTCACTACCCCTGGCTTCCCAACAGTCCCTACCCCGG		780			
	/db_xref="GI:13529566"									
	/db_xref="LocusID:19143"		Db	831	CCCATGGTGCAGCAGTGACACGCTTCACTACCCCTGGCTTCCCAACAGTCCCTACCCCGG		890			
	/translation="MGSNRRKAGGSGQDFGAGLKYNLSLENMNGFEVEFLPANNA									
	KXVEKRRPRWVLLVALEFSLLSLMAGLLVHHRVNRVQKVFNGHLRIITNEIFL		QY	781	CGCATGCCCGCTGCCAGTGGGCTCTCGCGGGGGAGCGCGGACTCTGTGCTGAGGCTCACCT		840			
	DAYENSTSEFTSLASQVKEAKLLYNEVPLGSPHKSAVAFSEGSVIATYWSFSS									
	IPPHLAERVDRAVAVRVTLLPPRARALKSVLTSVVAFPIDRMLQRTODNSCFAL		Db	891	CGCATGCCCGCTGCGCAGTGGGCTCTCGCGGGGGAGCGCGGACTCTGTGCTGAGGCTCACCT		950			
	HAGGAATRTFTPGPNSVPYAHARQWLRGDADSVLSLTPSDVAPCDSEHSDLV									
	TVYDLSLSPNEPHAVRLCGTSPSYNLTFLSSQNVLVTLINTDRHGFEPATFFQL		QY	841	TCCGAAGCTTTGATGTGCTCCCTCTGTGATGAGCATGGCAGTGCCTGGTCAACCGTGTATG		900			
	PMSSCGFLSDTQGFSPYPYGHYPNINCTWNKVNRRNRVRFKFLVFLVDNV									
	PVGSCTKDVEINGEKYGERBQFVSSNSKITVPHSDHSYDTGTGFLAEYLSYDSN		Db	951	TCCGAAGCTTTGATGTGCTCCCTCTGTGATGAGCATGGCAGTGCCTGGTCAACCGTGTATG		1010			
	DPCGMFKCTGRICRKLRCGDWADCPDYSDERYCENATHQFTCKNQFCKPLFWVC									
	DSVNDCCGDSDEGSCPCAGSPKSNKCLPOSCKNGKNDGDSDEASCDSDVNVVS		QY	901	ATAGCTGAGCCCCCATGGAACCCCAACCGCTGTGCTGGGCTGTGTGGACACTTCTCACCCCT		960			
	CTKYTYRCNGELCLSKNPECDGKTDPSGSEKNCDCGLRSFTQARVVGSTNADEG									
	ERPQVSLHALGQHLGASLISPDWLYSAHCFQDDRNFKYSDYMTWTFAGLLDQS		Db	1011	ATAGCTGAGCCCCCATGGAACCCCAACCGCTGTGCTGGGCTGTGTGGCACCTTCTCACCCCT		1070			
	KRSASGVQLKRLIITHSFNDFTEYDIALLELEKSVEYSYTVVRPILCPDATHVFP									
	AGKAIWVGHTKGGTGGALILQKGEIRVINTQCTEDLMPQOITPRMCMVGLSGGV		QY	961	CCTACAACTGACTTCTCTCTCCCAAGAACGCTCTTCTTGTTCAGCTGTATAACCAATA		1020			
	DSCQDSGGLPLSSAEXDGRMFQAGVSWEGSCAQRNKPVGVTLPVVRWDIKEHTGV"									
BASE COUNT	734 a 918 c 928 g 668 t		Db	1071	CCTACAACTGACTTCTCTCTCCCAAGAACGCTTCTTCTTGTACGCTGTATAACCAATA		1130			
ORIGIN										
	Query Match 99.9%; Score 3102.8; DB 10; Length 3248;		QY	1021	CTGACCGGCGACATCTCTGGCTTTGAGGCGACCTTTCTTCCAGCTGCCCAAGATGAGCAGCT		1080			
	Best Local Similarity 99.9%; Pred. No. 0;									
	Matches 3104; Conservative 0; Mismatches 2; Indels 0; Gaps 0;		Db	1131	CTGACCGGCGACATCTCTGGCTTTGAGGCGCACTTTCTTCCAGCTGCCCAAGATGAGCAGCT		1190			
			QY	1081	GTGGCGGCTTTTTCAGTGACACCCCAAGGGACATTTAGCAGCCCTTACTATCCAGGCCACT		1140			
			Db	1191	GTGGCGGCTTTTTCAGTGACACCCCAAGGGACATTTAGCAGCCCTTACTATCCAGGCCACT		1250			
			QY	1141	ACCCGCCCAACATCAACTGSCACATGGAATATCAAGGTGCCCAACACCGGAAACGTGAAGG		1200			
			Db	1251	ACCCGCCCAACATCAACTGSCACATGGAATATCAAGGTGCCCAACACCGGAAACGTGAAGG		1310			
			QY	1201	TGCGCTTCAAATCTTCTTATCTGTGGACCCCAAGTACCACTGGGCTCTCTGCAACCAAGG		1260			
			Db	1311	TGCGCTTCAAATCTTCTTATCTGTGGTGGACCCCAACGTCACAGTGGGCTCTCTGCAACCAAGG		1370			
			QY	1261	ACTATGTGAGCATCAACGGGGAGAAAGTACTGCGGTGAGAGGTCCTCAGTTTGTGGTGAGCA		1320			
			Db	1371	ACTATGTGAGCATCAACGGGGAGAAAGTACTGCGGTGAGAGGTCCTCAGTTTGTGGTGAGCA		1430			
			QY	1321	GCAACAGCAGCAAGATTACAGTCCACTTTCATTTCTGATCACTCGTACACGGAACACCGGGT		1380			
			Db	1431	GCAACAGCAGCAAGATTACAGTCCACTTTCATTTCTGATCACTCGTACACGGAACACCGGGT		1490			
			QY	1381	TCCTAGCTGAGTACTCTCTCTAGCTCCAAACGACCCGTCGCCAGGAGTGTTCATGTGCA		1440			
			Db	1491	TCCTAGCTGAGTACTCTCTCTAGCTCCAAACGACCCGTCGCCAGGAGTGTTCATGTGCA		1550			
			QY	1441	AGACTGGACGGTGCATCCGAAAGGAACTGCGCTGCGACCGGCTGGGAGACTGCCCGGATT		1500			
			Db	1551	AGACTGGACGGTGCATCCGAAAGGAACTGCGCTGCGACCGGCTGGGAGACTGCCCGGATT		1610			
			QY	1501	ATAGTGAATGAGGTTACTGTCCGATGCAATGCCAACCCAGTTTCACTGCAAAAAACCACT		1560			
			Db	1611	ATAGTGAATGAGGTTACTGTCCGATGCAATGCCAACCCAGTTTCACTGCAAAAAACCACT		1670			
			QY	1561	TCTGCAAGCCCTCTTCTTGGGTCTGTGACAGTGTCAACGACTGTGGGGAGCGGAAGTGAAG		1620			
			Db	1671	TCTGCAAGCCCTCTTCTTGGGTCTGTGACAGTGTCAACGACTGTGGGGAGCGGAAGTGAAG		1730			
			QY	1621	AGGAGGGCTGACAGTGTCTCTGTGGGAGTTTCAAGTGTTCGAATGGGAAGTGTCTCCCTC		1680			
			Db	1731	AGGAGGGCTGACAGTGTCTCTGTGGGAGTTTCAAGTGTTCGAATGGGAAGTGTCTCCCTC		1790			

Qy	2761	GCATCCTTAGAGACTCAGAGTTCTTCCMAAGTGGGACCCCTCAGAGTTGGAGAGAGAAC	2920
Db	2871	GCATCCTTAGAGACTCAGAGTTCTTCCMAAGTGGGACCCCTCAGAGTTGGAGAGAGAAC	2930
Qy	2821	TTGCGTGCTAGCGGCCCCAGCGCTGGGGGCAAGGTTTGTATGGCAGACCTTCCCTCTCTAGCC	2880
Db	2931	TTGCGTGCTAGCGGCCCCAGCGCTGGGGGCAAGGTTTGTATGGCAGACCTTCCCTCTCTAGCC	2990
Qy	2881	CTGAGCTGGGTGAAGATGATGCTGTGCCGAGAGCTGCTTCCAACTGTCTCATTCAGCTCCC	2940
Db	2991	CTGAGCTGGGTGAAGATGATGCTGTGCCGAGAGCTGCTTCCAACTGTCTCATTCAGCTCCC	3050
Qy	2941	GGGAGGCCATATGGAGAGAGGGGCTCAGGCTCAGCTCTTTTTCAGGAAGCGCCAGCCCTAGGA	3000
Db	3051	GGGAGGCCATATGGAGAGAGGGGCTCAGGCTCAGCTCTTTTTCAGGAAGCGCCAGCCCTAGGA	3110
Qy	3001	ACCCAGAAAGAGTGGTACCTAAGGCTGAAATTTGTTTGTCTTCCAGGGGTGGGTAT	3060
Db	3111	ACCCAGAAAGAGTGGTACCTAAGGCTGAAATTTGTTTGTCTTCCAGGGGTGGGTAT	3170
Qy	3061	TTGAGAGTAAACATTTTATTTCTTTTAAAAA	3106
Db	3171	TTGAGAGTAAACATTTTATTTCTTTAAAAA	3216
RESULT 4			
AB049189			
LOCUS			
DEFINITION			
Rattus norvegicus mRNA for membrane bound arginine specific serine			
protease, complete cds.			
ACCESSION			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
Rattus norvegicus (Norway rat)			
Rattus norvegicus			
Eukaryota; Metazoa; Chordata; Craniata, Vertebrata; Euteleostomi;			
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;			
Rattus.			
1 (sites)			
Inoue,H., Takahashi,K. and Kishi,K.			
membrane-bound arginine specific serine protease			
Published only in DataBase (2000)			
2 (bases 1 to 3174)			
Inoue,H., Takahashi,K. and Kishi,K.			
Direct Submission			
Submitted (22-SEP-2000) Hideshi Inoue, Tokyo University of Pharmacy			
and Life Science, School of Life Science, 1432-1 Horinouchi,			
Hachioji-shi, Tokyo 192-0392, Japan (E-mail:hinoe@ls.toyaku.ac.jp,			
Tel:81-426-76-7153, Fax:81-426-76-7157)			
Location/Qualifiers			
1. .3174			
/organism="Rattus norvegicus"			
/mol_type="mRNA"			
/strain="Wistar"			
/db_xref="taxon:10116"			
/sex="male"			
/tissue_type="duodenum"			
/dev_stage="7 weeks old"			
127. .2694			
/codon_start=1			
/evidence=experimental			
/product="membrane bound arginine specific serine			
protease"			
/protein_id="BAB13765.1"			
/db_xref="GI:10336527"			
/translation="MGNNRKGAGGSGD FGAGLYNGRLNMMNGFEEGVFLPVNNA			
KQVEKRGSTPRVVMVAVVSVFLLSLMAGLLVHHPYRNVR IQKVFNGHLRTNENFL			
DAVENSSTRTSLASQVEALKMYSEVPLGVPHKVTVTAFEGSGVIAVYWEPS			
IPPHLEEVDRAMAVRVVTVPPRARALKSFVLTSVAFVDPDPMQLQTDNSCSFAL			
HARGTVTRVTFPTGPNFNSPYPAHRCQWLVLRGDAVLSLSTRFDVAPCDGDSULV			
TVYDLSLSEPHAVVLCGTSFSPYNLTFLSQNVFLVTLITNDRRPPGEFATFQVL			
PKMSCKDGLYLSAQGTFFSPYPGPNNCTWNIKPVNNRNVKRFKFLVLDVPI			
PVGSCCTQVYEINGEKFCERSQFVVSNSKLTVHFHSDHSYTDTGFLAEVLSYDSN			
CDS			

DPCCGPMCKTRGIRKDLRCDGWADCPDYSDEHRCRNATHQFMCKNQFCKPLFWYC
DSVNDGDSDEBSCSPAGSFKNSGKCLPQSQCNKGDGDSDEASCDNVAVS
CTKYTRCQNGLCNKGNPECDGKKDCSDGSKNCDGLRSFTKQARVVGGTNADBG
EWPWSLHALGQHLGASLISPDWLVSAAHCFODETIFKYSNDHTWTAFLGLDOS
KBSASGVQEHKLEKLIHPSPNPTEDYDIALLELEXPABYSVTVRPICLPDNTHVPP
AGKAIWIGHYKGGTGAALLQKGEIRVINQTTCELLPQQTTPMMCVGELSGV
DSQGDSSGLPSSVEKDGRIFOAGVSWGSGCAORNPQVYTRPEVRDWIKESQTGV"
360
/note="compared with the sequence reported in Acc#
AB037898"
/replace="t"
513
variation
/note="compared with the sequence reported in Acc#
AB037898"
/replace="t"
573
variation
/note="compared with the sequence reported in Acc#
AB037898"
/replace="t"
2121
variation
/note="compared with the sequence reported in Acc#
AB037898"
/replace="t"
3157.3174
polyA_site
BASE COUNT 697 a 887 c 911 g 679 t
ORIGIN
Query Match 83.5%; Score 2595; DB 10; Length 3174;
Best Local Similarity 91.1%; Pred. No. 0;
Matches 2853; Conservative 0; Mismatches 230; Indels 50; Gaps 7;
QY 1 CATGGTAGCGGTGCGCGAGGAGCACCGTCTGAGACCGGCGATCGGACCGCCAAA 60
DB 65 CATGGTAGTTCGGTTCGGCGAGAGACACCGGTCTGAGACCGGCGATCGGCGCGCGA 124
QY 61 CCATGGGTAGCAATCGGGGCGCAAGCGCGAGGGGCTCTCAGACTTCGGCGGGGAC 120
DB 125 CCATGGGGAACAATCGGGGCGCAAGCGCGAGGGGCTCGCAGGACTTCGGCGCGGAC 184
QY 121 TCAAGTCAACTCCCGGTAGAGAACATGAATGCTTTGAGAGGGGTGTGAGTTCCTGC 180
DB 185 TCAAGTCAACTCCCGGTAGAGAACATGAATGCTTTGAGAGGGGTGTGAGTTCCTGC 244
QY 181 CTGGGAACAATGCCAAGAATGAGAGAGAGCGCCCGAGCGGTGCGGTGTGTTGGTGG 240
DB 245 CTGTGAACAACGCCAAGCAGGTGAGAGAGAGCGCCCGAGCGGTGCGGTGTGTTGGTGG 304
QY 241 CAGTGTCTTTCAGCTTCTCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 300
DB 305 CCGTAGTGTTCAGCTTCTCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 364
QY 301 ATTATCGGAATGTGCGGGTTCAAAAAGTCTTCAATGGCCATCTGAGGATCACAATGAGA 360
DB 365 ATTACCGGAATGTGAGGATCCAAAAGTCTTCAATGGCCATCTGAGGATCAAGATGAGA 424
QY 361 TCTTTCTGGATGCGTATGAGAACTCCACCTCCACAGAGTTTATCAGCTGCGCAGCAGG 420
DB 425 ACTTCTTGGACGCGTATGAGAACTCCACCTCCACGAGTTTCATCAGCTGCGCAGTCA 484
QY 421 TGAAGAGGCGCTGAAGCTGTGTACAAATGAAGTCCCTGCTGCTGCTGCTGCTGCTGCT 480
DB 485 TGAAGAGGCGCTGAAGCTGTGTACAAATGAAGTCCCTGCTGCTGCTGCTGCTGCTGCT 544
QY 481 AGTCCGCTGTAACTGCTTCACTGAGGCGAGTGTCACTCGCTACTACTGTGTCAAGTTCA 540
DB 545 AATCGACTGTAACTGCTTCACTGAGGCGAGTGTCACTCGCTACTACTGTGTCAAGTTCA 604
QY 541 GCATCCCCCACCTCGCAGAGAGGTTGATCGGCGCATGGCTGTGAGCGAGTTGTAA 600
DB 605 GCATCCCCCACCTCGCAGAGAGGTTGATCGGCGCATGGCTGTGAGCGAGTTGTCA 664
QY 601 CATTGCCACCCGAGACGCGSCACTGAAATCCTTCGTGCTAAACATCTGTGTCCTTCC 660

665 CACTGCCACCCGAGACGCGGCACAGAAATCTCTGCTGCTGACCTCTGTGTGTCCTTCC 724
QY 661 CCATTGACCCCAAGTCTGACAGAGACTCAGAGAACAGCTGCAGTTTGTGTCCTGATG 720
DB 725 CCATTGACCCCAAGTCTCAGAGACTCAGAGAACAGCTGCAGTTTGTGTCCTTCCATG 784
QY 721 CCATGCTGACGAGTGAACAGCTTCACTACCCCTGCTGCTTCCCAACAGTCCCTTACC 780
DB 785 CCGTGGCAGGACAGTGACAGTTTCACTACCCCTGCTTCCCTTAAACAGTCCCTTACC 844
QY 781 CGATGCGCGCTGCGAGTGGTCTTGGGGGGAGCGGAGCTGTGCTGAGCTTCACTT 840
DB 845 CTATGCGCGCTGCGAGTGGTCTTGGGGGGAGCGGAGCTGTGCTGAGCTTCACTT 904
QY 841 TCCGAAGCTTTGATGTGCTGCTGCTGATGAGCATGGCAGTGCCTGCTGCTGCTGCTG 900
DB 905 TCCGAAGCTTTGATGTGCTGCTGCTGATGAGCATGGCAGTGCCTGCTGCTGCTGCTG 964
QY 901 ATAGCTGAGCCCATGGAAACCCACGCTGTGTGCGGCTGTGTGGCACCTTCTTCACT 960
DB 965 ACAGCTGAGCCCATGGAGCCCATGCTGTGTGCGGCTGTGTGGCACCTTCTTCCGCT 1024
QY 961 CTTACAACTGACTTCTCTCTCTCCAGAACGCTTCTTCTGCTGCTGCTGCTGCTGCTG 1020
DB 1025 CTTACAACTGACTTCTCTCTCTCCAGAACGCTTCTTCTGCTGCTGCTGCTGCTGCT 1084
QY 1021 CTGACCGGCGACATCTCTGCTTGGAGCCACTTCTTCTTCCAGCTGCCCCAAGATGAG 1080
DB 1085 CTGACCGGCGACATCTCTGCTTGGAGCCACTTCTTCTTCCAGCTGCCCCAAGATGAG 1144
QY 1081 GTGGCGCTTTTGTGAGTGACACCCAAAGGAGCAATTTAGACGCCCCCTACTATCCAG 1140
DB 1145 GTGGCGCTTTTGTGAGTGAGCCAAAGGAGCAATTTAGACGCCCCCTACTATCCAG 1204
QY 1141 ACCGCGCCAACTCAACTGACATGCAATGAAATCAAGTGGCCCAACACCGGAACTGAG 1200
DB 1205 ACCGCGCCAACTCAACTGACATGCAATGAAATCAAGTGGCCCAACACCGGAACTGAG 1264
QY 1201 TGCGCTTCAAACCTCTTCTATCTGCTGAGACCCCAACGCTTACAGTGGGCTCTGCA 1260
DB 1265 TACGCTTCAAACCTCTTCTATCTGCTGAGACCCCAACATACAGTGGGAACTGCA 1324
QY 1261 ACTATGTGAGATCAACCGGAGAAAGTATGCGGTGAGAGTCCAGTTTGTGTTGAGCA 1320
DB 1325 ACTATGTGAGATCAACCGGAGAAAGTATGCGGTGAGAGTCCAGTTTGTGTTGAGCA 1384
QY 1321 GCACAGCAGCAAGATTTACAGTCCACTTCCATTTCTGATCACTGTAACGACACCG 1380
DB 1385 GCACAGCAGCAAGATTTACAGTCCACTTCCATTTCTGATCACTGTAACGACACCG 1444
QY 1381 TCTAGCTGAGTACCTCTCTTACGACTCCAAACGACCCGCTGCCCCAGGAGTGTTCAT 1440
DB 1445 TCTGCTGAGTATCT 1504
QY 1441 AGACTGAGCGTGCATCCGAAAGAACTGGGCTGCGACCGCTGGGAGACTGCCCCGAT 1500
DB 1505 AGACTGAGCGTGCATCCGCAAGAACTTGGCTGCGATGGCTGGGCTGCTGCCCCG 1564
QY 1501 ATAGTATGAGCGTGTACTGCGGATGCAATGCCCCACCACTGCTGCAAAAACAGT 1560
DB 1565 ACAGTATGAGCGTGTACTGCGGATGCAACGTTACCCACCACTGTTCTGTAAGAAC 1624
QY 1561 TCTGCAAGCCCTCTTCTGCTGCTGAGCTGTGACAGTGTCAACGCTGTTGGGAGCG 1620
DB 1625 TCTGCAAGCCCTCTTCTGCTGCTGAGCTGTGACAGTGTCAACGCTGTTGGGAG 1684
QY 1621 AGAGGGCTGAGCTGCTGCTGAGGTTTCAAGTGTTCATGAGGAGTGTCTCTCT 1680
DB 1685 AGAGGGCTGAGCTGCTGCTGAGGTTTCAAGTGTTCATGAGGAGTGTCTCTCT 1744
QY 1681 AGAGCCAGAGTGTCTAATGGGAAGGACAACTGTGAGATGGTCTGACGAGGCTTCA 1740
DB 1745 AGAGCCAGAGTGTCTAATGGGAAGGACAACTGTGAGATGGTCTGACGAGGCTTCA 1804

QY 1741 ACAGCGTGAATGTCGTCTCTCTGACAAATATACCTACCGCTGCCAAATGCGCTCTCTC 1800
Db 1805 ACAATGTGAATGCGGTCTCTCTGACAAATATACCTACCGCTGCCAAATGCGCTCTCTC 1864
QY 1801 TGAGCAAGGGCAACCTCAGTGTGTGATGGAGACGGAAGCTGTAGCGATGCTCCGATGAGA 1860
Db 1865 TGAACAAGGGCAACCTCAGTGTGTGATGGAGAGGAGACTGTCAGCGATGGCTCGGATGAGA 1924
QY 1861 AAAAATGTGACTGTGGGTGGGATCTTTTACCAACAGGCTCGCGTGGTGTGGTGACAGA 1920
Db 1925 AAAAATGTGACTGTGGGTGGGATCTTTTACCAACAGGCTCGAGTGGTGTGGTGACAGA 1984
QY 1921 ATGGGACAGAGGGGAGTGGGCCCTGCGAGCTGAGCCTCCACGCCCTGGGGCCAGGSCCACT 1980
Db 1985 ATGGGACAGAGGGGAGTGGGCCCTGCGAGCTGAGCCTCCACGCCCTGGGGCCAGGSCCACT 2044
QY 1981 TGTGTGGGGCTCGCTCATCTCTCTGACTGGTGGTGTCTCTGAGCTCATTTGCTTTTACGG 2040
Db 2045 TGTGTGGGGCTCGCTCATCTCTCTGACTGGTGGTGTCTCTGAGCTCATTTGCTTTTACGG 2104
QY 2041 ATGACAAAATTTCAAGTACTCAGACTACAGATGTGAGCGGCTTCTGCTGCTCTGCTGG 2100
Db 2105 ATGACAAATTTTCAAGTACTCAGACTACAGATGTGAGCGGCTTCTGCTGCTCTGCTGG 2164
QY 2101 ACCAGAGCAAGCGAGTGCCTCTGGGGTGCGAGGCTGAAGCTCAAAAGCTATCATCAACC 2160
Db 2165 ACCAGAGCAAGCGAGTGCCTCTGGGGTGCGAGGCTGAAGCTCAAAAGCTATCATCAACC 2224
QY 2161 ACCCTCTCTCAATGATTTACCTTCGACTATGATCATCGCCTTGTGAGCTGAGAGT 2220
Db 2225 ACCCTCTCTCAATGATTTACCTTCGACTATGATCATCGCCTTGTGAGCTGAGAGT 2284
QY 2221 CGGTGGAGTACAGCACCTCTGTGCGCCCATCTGCTGCTGATGCTACCCATGCTTCC 2280
Db 2285 CGGAGAGTACAGCACTGTGTGCGCCCATCTGCTGCTGATGATTAACCATGCTTTC 2344
QY 2281 CTGTGGCAAGGCCATCTGGGTCTCAGCTGGGGCACACAAAAGAGGGAGGTACCGGAG 2340
Db 2345 CTGCGCGCAAGGCCATCTGGGTCTCAGCTGGGGCACACAAAAGAGGGAGGTACCGGAG 2404
QY 2341 CGCTGATCTCTCAGAGGGTGAATCCCTGTGTCATCAACAGCACACCTGTGAGACCTCA 2400
Db 2405 CACTGATCTCTCAGAGGGAGAGATCCGGGTCTATCAACAGCACACCTGTGAGAGTCT 2464
QY 2401 TCGCGCAGCAGATCACCCACGAATGATGTGTGGGTTTCTCAGTGGGGGTGTGACT 2460
Db 2465 TCGCGCAGCAGATCACCCACGAATGATGTGTGGGTTTCTCAGTGGGGGTGTGACT 2524
QY 2461 CCGCGCAGGGTACTCTGGTGGCCCTTTGTCAAGCGCGGAGAAAGATGGCGAATGTTTC 2520
Db 2525 CCGCGCAGGGTACTCTGGTGGCCCTTTGTCAAGCGCGGAGAAAGATGGCGAATGTTTC 2584
QY 2521 AGGCTGTGTGTGAGCTGGAGTGAAGCTGGCTCAGAGGACAAAGCAGCGCTGTACA 2580
Db 2585 AGGCTGTGTGTGAGCTGGAGTGAAGCTGGCTCAGAGGACAAAGCAGCGCTGTACA 2644
QY 2581 CAAGGCTCCCTGTAGTTCGGGACTGGATCAAGAGCACACTGGGGTATAGCAGATGAC 2640
Db 2645 CAAGGATCCCTGAAGTTCGGGACTGGATCAAGAGCAAACTGGGGTATAGAGCATGGAC 2704
QY 2641 AGACAGCGGACACAAACACCCACAGGATGCCGACATGCACACCTGGATACAGGAGAG 2700
Db 2705 AGACAAACCCACCATGAATCTCGAAAGGGATGCCCGCCACCGGCTACCTGGATACAGCAG 2764
QY 2701 GAACACTAGCAGCATTTATGCTGTGGCTTCCCCCCCCCAACACACCCAGACTGTGACT 2760
Db 2765 GATCGCTAGGACATTTGTGTGTGGCTCTCCCC-----CACATCCAGAGCTGTGACT 2819
QY 2761 GCATCCCTTAGGACTCAGAGTTCTTCCA-----AAGTGGGACCCCTCAAGAGTTGG 2811
Db 2820 GCTTCCCTTAGGACTCAGAGTTCTTCCCTCCCTGAGTTCCAAAGTGGGACCCCTCAGAGTTGG 2879

QY 2812 AGAGAGAACTTGCGTGTAGCGGCCAGCCCTG-----GGGCAAGGGTTTGTAGCGAGCT 2867
Db 2880 CCAGAGATCTTCTGTGTGTGGCCAGCCCGGGTGGGCGAGGATTTGATGGCGCT 2939
QY 2868 TCCCCCTCTAGCCCTGAGCTGGGTGAAGATGATCTGTCCCGGAGAGCTGCTTC----- 2922
Db 2940 T-CCCTCTTAACTTGAACCTGGTAAAGACGATGCTTCTCTGGAGAACTGCTTCCAGACT 2998
QY 2923 -----AACTCTCATTTGAGCTCCCGGAGCCCTATGGGAGGAGGGCTCAGGCTAC 2973
Db 2999 CTAAGGAAATATATGCTGCGCTCTCTGGGGCCCTATGGGCGAGAGGGCTGCGAG----- 3054
QY 2974 TCTTTTCAGGAAGCGCCAGCCCTAGGAACCCAGAAAAGAGTGGTACCTAAGGCTGAAT 3033
Db 3055 GTCTTCAGGAAGCTCCAGCCCTAGGACCCAGAAAAGAGTGGCACCTATGGCTGATT 3114
QY 3034 TGTTCCTGTGTGCGAGGGTGGGTATTTGAGAGTAAACATTTTATTTCTTTTAAAAA 3093
Db 3115 TGTTCCTGT-----GGTATTTGAGAGTAAACATTTTATTTCTTTTAAAAA 3161
QY 3094 AAAAAAATAAAAA 3106
Db 3162 AAAAAAATAAAAA 3174

RESULT 5
AB037898
LOCUS
DEFINITION
AB037898 2568 bp mRNA linear ROD 20-OCT-2001
Rattus norvegicus MBSP mRNA for membrane bound serine protease,
complete cds.
ACCESSION
AB037898
VERSION
AB037898.1 GI:9650963
KEYWORDS
membrane bound serine protease; MT-SP1.
SOURCE
Rattus norvegicus
ORGANISM
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE
1
AUTHORS
Satomi,S., Yamazaki,Y., Tsuzuki,S., Hitomi,Y., Iwanaga,T. and
Fushiki,T.
TITLE
A role for membrane-type serine protease (MT-SP1) in intestinal
epithelial turnover
JOURNAL
Biochem. Biophys. Res. Commun. 287 (4), 995-1002 (2001)
MEDLINE
21458307
PUBMED
11573963
REFERENCE
2 (bases 1 to 2568)
AUTHORS
Tsuzuki,S.
TITLE
Direct Submission
JOURNAL
Submitted (26-JAN-2000) Satoshi Tsuzuki, Kyoto University, Graduate
School of Agriculture, Division of Food Science and Biotechnology,
Laboratory of Nutrition Chemistry; Oiwake-cho, Kitashirakawa,
Sakyo-ku, Kyoto, Kyoto 606-8502, Japan
(E-mail:tkmone@kais.kais.kyoto-u.ac.jp, Tel:81-75-753-6263,
Fax:81-75-753-6264)
FEATURES
source
1. .2568
/organism="Rattus norvegicus"
/mol_type="mRNA"
/strain="Wistar"
/db_xref="taxon:10116"
/sex="male"
/tissue_type="jejunum"
/dev_stage="8 weeks old"
1. .2568
/gene="MBSP"
1. .2568
/gene="MBSP"
/codon_start=1
/product="membrane bound serine protease"
/protein_id="BAB03502.1"
/db_xref="GI:9650964"
/translation="MGNNRGRKAGGSGQDFGAGLYKNSRLNMMNMFEGVEFLPVNNA"

BASE COUNT
ORIGIN

559 a	708 c	746 g	555 b

Query Match 74.5%; Score 2313.6; DB 10; Length 2568;

Best Local Similarity 93.8%;
Matches 2409; Conservative 0; Mismatches 159; Indels 0; Gaps 0;

63	ATGGGTAGCAATCGGGGCCGCAAGCCGAGAGGGGCTCTCAGACATTCGGCGCGGAGCTC	121
QY		
1	ATGGGGAACAAATCGGGGCGCAAGCCGAGAGGGGCTCGCAGGACTTCGGCGCGGAGCTC	60
Db		
123	AAGTACAACTCCCGCTAGAGAACATGAATGGCTTTTCAGGAGGGTGTGGAGTTCCTGCT	182
QY		
61	AAGTACAACTCCCGCTAGAGAACATGAATGGCTTTTCAGGAGGGTGTGGAGTTCCTGCT	120
Db		
183	CGCAACAATCCAAAGAAGTGGAGAAGCGAGGCCCCAGGCGCTGGGCTGGTGTGGTGGCA	242
QY		
121	GTGAACAACGCCAAGCAGGTGGAGAAGGAGGCCCCAGGCGCTGGTGTGGTGGCC	180
Db		
243	GTGCTGTTCACTTCCTCTCTCTCTCCCTCATGCGCTGGCTTGTGTTGGCACTTCAT	302
QY		
181	GTAGTGTTCAGTCTTCCTTTTCTCTCTCATGGCCGGCTTGTGGTGTGGCACTTCAT	240
Db		
303	TATCGGAATGCGGGTTCAAAAGTCTCAATGGCCATCTGAGGATCACAATGAGATC	362
QY		
241	TACCGGAATGTAGAGATCCAAAAGTCTTCAATGGCCATCTGAGGATCAGGAATCAGAAC	300
Db		
363	TTTCTGGATCGGTATGAGAACTCCACCTCCACACAGATTTATCAGCTGCGCCAGCCAGGTG	422
QY		
301	TTTCTGGACGCGTATGAGAACTCCACCTCCACGAGTTTCAATCAGCTGCGCCAGTT	360
Db		
423	AAGGAGGCGCTGAAGTCTGTGTAAATGAAGTCCCTGTCTGGGTCCCTTACCACAAGAG	482
QY		
361	AAGGAAGCGCTGAAGTCTGTGTAAATGAAGTCCCTGTCTGGGTCCCTTACCACAAGAAA	420
Db		
483	TCGGCTGTAACTGCTCTCAGTGAGGGCAGTGTATCGCTACTACTGTCAGAGTTCAGC	542
QY		
421	TCGACTGTAACTGCTCTCAGTGAGGGCAGCGTATTGCCCTACTACTGTGTCAGAGTTCAGC	480
Db		
543	ATCCCCCACACCTGGCAGAGAGGTTGATCGCGCCATGSGCTGTGAGCGAGTTGTAAACA	602
QY		
481	ATCCCCCACACCTGGAGGAAGAGTTGATCGTGCCATGSGCTGTGAGCGAGTTGTCAACA	540
Db		
603	TGCGCACCCCGAGCACGGGCATGAAATCTCTGTGCTAAACATCTGTGTGGTGGCTTCCCC	662
QY		
541	CTGCCACCCCGAGCACGGGCATGAAAATCTCTTTGTGCTGACCTCTGTGTGGTGGCTTCCCC	600
Db		
663	ATTGACCCCAAGATGCTGCAGAGACATCAGACAACAGCTGCAGTTTTTGCCTGTGATGCC	722
QY		
601	ATTGACCCCAAGATGCTCCAGAGACATCAGACAACAGCTGCAGTTTTTGCCTTCCATGCC	560
Db		
723	CATGTGCAGCAGTGACACGCTTCACTACCCCTGGCTTCCCCAACAGTCCCTACCCGGCG	782
QY		
661	CGTGCAGGACAGTGACACGTTTCACTACCCCTGGCTTCCCTTAAACAGTCCCTACCCAGCT	720
Db		
783	CATGCCCGCTGCAGTGGGTCTCGGGGGGGAAGCGCGACATCTGTGTGAGCCTCACCTTC	842
QY		
721	CATGCCCGCTGCAGTGGGTCTCGGGGGGGAAGCGGACATCTGTGTGAGCCTCACCTTC	780
Db		

743	Db	CCCCCTACCCCGCTCATGCCCCGCTGCCAGTGGCGCCTTCGGGGGAGCCGCACCTCACTGCT	802
830	QY	GAGCCTCACCTTCCGAAGCTTTTGATGTGCGTCCCTGTGATGAGCATGCGAGTGACCTGGT	889
803	Db	GAGCCTCACCTTCCGAGCTTTGACCTTGCTCTGCGACGAGCGCGCAGACCTGGT	862
890	QY	CACCGTGATGATAGCTGAGCCCCATGGAACCCACGCTGTGGTGGCGCTGTGTGGCAC	949
863	Db	GACGGTGTAACAACCTGAGCCCCATGGAAGCCACGCGCTGGTCAGTTGTGTGGCAC	922
950	QY	CTTCTCACCTCTCAACAACCTGACTTTCTCTCTCCAGAACGTCTTCTGTGTGACGCT	1009
923	Db	CTACCTCTCCTCTCAACAACCTGACTTCCACTCTCCAGAAAGTCTGCTCATCACACT	982
1010	QY	GATAACCAATATGACCCGGGCAATCCTGTGCTTTGAGGCCACATTTCTTCAGTGGCCAA	1069
983	Db	GATAACCAACACTGAGCGCGGCGATCCGCGCTTTGAGGCCACCTTTCTTCAGTGGCTAG	1042
1070	QY	GATGAGCAGCTGTGGCGGCTTTTGTAGTGACCCCAAGGGACATTTAGCAGGCCCTACTA	1129
1043	Db	GATGAGCAGCTGTGGAGCGCTTACGTAAAGCCAGGGGACATTTCAAGACCCCTACTA	1102
1130	QY	TCCAGGCGCACTACCCGCCAACAATCACTGCAATGGAATATCAAGGTGCCCAACAACCG	1189
1103	Db	CCAGGCGCACTACCCACCACCAATTGACTGCAATGGAAATTTGAGGTGCCCAACAACCA	1162
1190	QY	GAACTGAAAGTGCCTTCAAACTCTTCTATCTGTGTGACCCCAAGTACCATGTGGGCTC	1249
1163	Db	GCATGTGAAAGTGCCTTCAAACTCTTCTATCTGTGTGAGCCCGGCTCTGGGGCAC	1222
1250	QY	CTGCACCAAGGACTATGTGGAGATCAACGGGGAGAAAGTATCTGGCGGTGAGAGGTCCAGTT	1309
1223	Db	CTGCCCCAAGGACTATGTGGAGATCAATGGGGAGAAATCTGCGGAGAGAGGTCCAGTT	1282
1310	QY	TGTGTTGACGACCAAGCAGCAGCAAGATATACAGTCACTTCCATTTGATCACTCGTACAC	1369
1283	Db	CGTGTGTACACAGCAACAGCAACAAAGATCAAGTTGCGTTCCACTCGATCAAGTCTTACAC	1342
1370	QY	GGACACCGGTTTCTTAGCTGAGTACCTCTCTTACGACTCCAAAGCCCGTGCCACGAGAT	1429
1343	Db	CGACACCGGTTTCTTAGCTGAATACCTCTCTTACGACTCCAGTGAACCATGCCCGGGCA	1402
1430	QY	GTTTATGTGACAGCTGAGCGGTGCATCCGAAGGAATCTGCGTGTGGAAGCGCTGGGCGA	1489
1403	Db	GTTTACGTGCGCACCGGGCGGTGTATCCGGGAAGAGCTGCGTGTGATGGCTGGGCGGA	1462
1490	QY	CTGCCCAGGATTAATGATGAGCGTTACTGSCCGATGCAATGCCACCCACAGTTTCAGTG	1549
1463	Db	CTGCACCGACCAAGCGATGAGTCACTGCACTTGGACCGCGGCCACAGTTTCAGTG	1522
1550	QY	CAAAAACCAAGTTCTCAAGCCCTCTTCTGGGTCTGTGACAGTGTCAACAGCTGTGGGGA	1609
1523	Db	CAAGAAACAAGTTCTGCAAGCCCTCTTCTGGGTCTGCGACAGTGTCAACAGCTGTGGGGA	1582
1610	QY	CGGAAGTGAAGGAGGCGTCAAGCTGTCTGCTGGAGTTTCAAGTGTTCGAATGGGAA	1669
1583	Db	CAACACGACGACGAGGGGTGCAAGTGTGCGGCCCAAGCTTCAAGTGTGTCCAAATGGGAA	1642
1670	QY	GTGTCTCCCTCAGAGCCAGAAGTGTAAATGGGAAGGACAACTGTGGAGATGGGTCTGACGA	1729
1643	Db	GTGCTCTCGAAAAGCCAGCAGTGCATGGGAAGGACGACTGTGGGGAAGCGGTCCGACGA	1702
1730	QY	GGCTTATGTGACAGCGTGAATGTGCTCTTTCGACCAAAATATACCTACCGCTGCCAAA	1789
1703	Db	GGCTCTCTGCCCAAGGTGAACGTGTGCTCACTTGTACCAAAACACACCTTACCGCTGCCCTCA	1762
1790	QY	TGGCTCTGTCTCAGCAAGGCAACCTGAGTGTGATGGGAAGCAGGACGTGAGCGATGG	1849
1763	Db	TGGCTCTGTCTGAGCAAGGCAACCTGAGTGTGACGGGAAGGACGTGTGAGCGACGG	1822
1850	QY	CTCCGATGAAAAAATGTGACTGTGGGCTGCGATCTTTTACCAAAACAGCGCTCGCGTGT	1909

Db	1823	CTCAGATGAGAAAGGACTGGCACTGTGGCTGGCGGTCAITTCACGAGACAGGCTCGTCTTGT	1888
Qy	1910	TGCTGGCAGAAATGCGGACGAGGCGAGTGGCCCTGGCAGGTGAGCTCCACGCCCTGGG	1969
Db	1883	TGGGGGCACGGATGCGGATGAGGGCGAGTGGCCCTGGCAGGTAAAGCTGATGCTCTGGG	1942
Qy	1970	CGAGGGCACTTGTGTGTGGGCTTCGCTCACTCTCTCTGACTGGCTGGTCTCTGCACTCA	2029
Db	1943	CCAGGGCCACATCTGGGTGCTTCCCTCATCTCTCCAACTGGCTGGTCTCTGCGGCACA	2002
Qy	2030	TTGCTTTCCAGATGACAAAATTTCAAGTACTCAGACTACAGATGTGAGCGGCTTCCT	2089
Db	2003	CTGCTACATCGATGACAGAGGATTCAGGTACTCAGACCCACGCACTGAGCGGCTTCCT	2062
Qy	2090	GGGTCTGTGGACACAGACGAGCGCAGTGCCTCTGGGGTGCAGGAGCTGAAGCTCAACG	2149
Db	2063	GGGCTTTCACGACACAGAGCCAGCGCAGCGCCCTGGGGTGCAGGAGCGCAGGCTCAAGG	2122
Qy	2150	TATCATCACCCACCCCTTCCTTCAATGATTTCACTTCGACTATGACATCGCTTGCTGGA	2209
Db	2123	CATCATCTCCACCCCTTCTTCAATGACTTCACCTTCGACTATGACATCGGCTGCTGGA	2182
Qy	2210	GCTTGGAGATGGTGGAGTACAGCACCGTCTGGGCCCCCATCTGCCTGCCCTGATCTAC	2269
Db	2183	GCTTGGAGAAACCGGCACAGATACAGCTCCATGTGTGGGCCCATCTGCCTCCGCGACGCTC	2242
Qy	2270	CCATGCTTCCCTGCTGGCAAGGCCCATCTGGGTCAAGCTGGGGGCACACAAAGAGGG	2329
Db	2243	CCAATGCTTCCCTGCGGCAGAGCCATCTGGGTACGGGCTGGGGACACACCCAGTATGG	2302
Qy	2330	AGGTACCGGAGCGCTGATCCTTCAGAAGGTGAGATCCGTTGTCATCAACACGACCACTTG	2389
Db	2303	AGGCACTGGCGGCTGATCCTGCAAAAGGTGAGATCCGCGTCACTCAACACGACCACTG	2362
Qy	2390	TGAGGACCTCATGCCGACGAGATCACCCCACGAATGATGTGTGGGTTTCTCAGTGG	2449
Db	2363	CGAAGACCTCTCCGCGACGAGATCACGCGCGATGATGCGTGGGGTTCCTCAGCGG	2422
Qy	2450	GGGTGTGGACTCTGTCCAGGGTGACTCTGGTGGCCCTTTGTCAAGCGCGGAGAAAGATGG	2509
Db	2423	CGGCGTGGACTCTGTCAGGGTGATTTCCGGGGGACCCCTGTCCAGCTGAGCGGATGG	2482
Qy	2510	GCAGATGTTCCAGGCTGTGTGTGAGCTGGGGTGAAAGGCTGGCTCAGAGGAAACAGCC	2569
Db	2483	GCGGATCTTCCAGGCCGCTGTGTGTGAGCTGGGGAGACGCTCGCTCAGAGGAAACAGCC	2542
Qy	2570	AGGCGGTGATACAAAGGCTCCCTGTAGTTCGGGACTGGATCAAGAGACACTGGGGTATA	2629
Db	2543	AGGCGTGTACADAGGCTCCCTCTGTTTCGGGACTGGATCAAGAGAACTCTGGGGTATA	2602
Qy	2630	GCAGCATGACAGACAGCCGACCAAAACCCACAGGATGCCCGACATGACACCTGG	2689
Db	2603	GGGGCCGGG---GCCACCCAAATGTGTACACCTGGGGGGCCACCCTCGTCCACCCCACT	2659
Qy	2690	ATACAGAGAGGAGAACTACGACATTTATGTGTGGCTCCCTCCCTCCCAACACACACCA	2749
Db	2660	GTGCACG-CCTGAGGCTGGAGCTGAGCCCGCTGACTGCACAGGCGCC-CCAGAACATA	2717
Qy	2750	GACTGTGAAGTGCATCCTTAGGACTCAGAGT	2780
Db	2718	CAGTGTGAAGTCAATCTCCAGGGCTCAAAAT	2748

RESULT 8
AF133086
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

AF133086 3120 bp mRNA linear PRI 01-OCT-1999
Homo sapiens membrane-type serine protease 1 mRNA, complete cds.
AF133086
AF133086.1 GI:6002713

Homo sapiens (human)
Homo sapiens


```
1474 GCGACGCTGGCAGACTGCCGGAATATATAGTGATGAGCGTTACTGCGGATGCAATGCCA 1533
1448 GTGATGGCTGGCGCAGCTGCACCGACACACGAGATGAGTCAACTGAGTTGCGACGCG 1507
1534 CCACACAGTTACGTGCAAAACAGTTCTGCAAGCCCTCTTCTGGGTCTGTGAAGTG 1593
1508 GCCACCAAGTTACGTGCAAGAACAGTTCTGCAAGCCCTCTTCTGGGTCTGTGACAGTG 1567
1594 TCAACGACTGTGGGACGGAAGTGCAGGAGGCGCTCAGCTGTCCTGTGGAGTTTCA 1653
1568 TGAACGACTGGGAGACAAACAGCAGCAGCAGGAGGTCAGTTGTCCGCGCCAGACCTTCA 1627
1654 AGTGTTCATAGGAAGTGTCTCCCTCAGAGCCAGAAAGTGTAAATGGGAAGCAACTGTG 1713
1628 GGTGTTCATAGGAAGTGTCTTCGAAAGCCAGCAGTGAATGGGAAGCAGCTGTG 1687
1714 GAGATGGTCTGACGAGGCTTCATGTACACGCTGAATGTGCTCTTTGACCAATATTA 1773
1688 GGGACGGGTCCGACGAGGCTCTCTGCCCAAGGTGAACGTCGTCACTTGTACCAACACA 1747
1774 CCTACCGCTGCCAAATGGCTCTGTCTGAGCAAGGCAACCTCAGCTGTGATGGGAAGA 1833
1748 CTTACCGCTGCCATAGGGCTCTGTCTGAGCAAGGCAACCTCAGTGTGACGGAGG 1807
1834 CGACTGTAGGATGGCTCCGATGAGAAACCTGTGATGTGGCTGCGATCCTTTACCA 1893
1808 AGGACTGTAGCAGGCTCAGATGAGAGGACTCGCACTGTGGCTGCGCTCATTCACGA 1867
1894 AACAGGCTCCGCTGTGTGGCAGCAATCGCAGAGGCGAGTGCCTGCGAGGTGA 1953
1868 GACAGGCTCTGTGTGGGGCACGATCGGATGAGGCGGAGTGGCCCTGGCAGGTAA 1927
1954 GCTTCCAGCCCTGGGCCAGGGCCACTGTGTGGGGCTCGCTCATCTCTCTGACTGGC 2013
1928 GCCTGCATGTCTGGGCCAGGGCCACATCTGCGGTGCTTCCCTCATCTCTCCCACTGGC 1987
2014 TGGTCTCTGACGCTCATGTCTTTCAGATGACAAAATTTCAAGTACTCAGATACAGA 2073
1988 TGGTCTCTGCGCAGACACTGTATCATCATGACAGAGGATTCAGGTACTCAGACCCACGC 2047
2074 TGTGAGCGGCTTCTCGGTCTGTGACACAGAGCAGCGCAGTGCCTCTGGGGTGCGAG 2133
2048 AGTGGAGCGCTTCTGGGCTTGCACACAGAGCAGCGCCCTGGGGTGCGAG 2107
2134 AGTGAGCTCAAAAGCTATCATCAACCCCTTCTTCAATGATTTTCACTTCGACTATG 2193
2108 AGCGAGGCTCAAGCGCATCATCTCCACCCCTTCTTCAATGACTTCACTTCGACTATG 2167
2194 ACATCGCTTGTGGAGCTGGAGAGTCCGTGGAGTACAGACCGTGTGCGGCCCATCT 2253
2168 ACATCGCTTGTGGAGCTGGAGAAACCGGACAGATACAGCTCCATGGTGCGCCCATCT 2227
2254 GCCTGCTGTATGCTACCCATGTCTTCCCTGCTGGCAGGCGCATCTGGGTCAACAGGTGGG 2313
2228 GCCTGCGGAGCGCTCCCATGTCTTCCCTGCGGCAAGGCCATCTGGGTCAAGGGTGGG 2287
2314 GGCACACAAAGAGGAGTACCGAGCGCTGTATCTCTGACAGAGGTGAGATCCGTGTCA 2373
2288 GACACACCCAGTATGGAGGCACTGGCGCGCTGATCTCTGCAAAAGGCTGAGATCCGGTCA 2347
2374 TCAACAGACCACTGTGAGGACCTCATGCGGAGCAGATCAACCCACGATATGTGTG 2433
2348 TCAACAGACCACTGGAGAACTCTCTGCGGAGCAGATCAACCCGCGCATATGTGCG 2407
2434 TGGGTTTCTCAGTGGGGGTGTGACTCTCTCCAGGGTGTACTCTGTGGGCCCTTGTCAA 2493
2408 TGGGTTCTCAGCGCGGCTGGAATCTCTCCAGGGTGTATTCGGGGGACCCCTGTCCA 2467
2494 GCGGGAAGAGTGGGCGAATGTTCCAGGCTGTTGAGTGTGAGTGGGGTGAAGGCTGGG 2553
2468 GCCTGGAGGCGGATGGGCGGATCTTCCAGGCGGTTGTTGAGTGGGAGACGGCTGGG 2527
2554 CTCAGGAGAACAGCCAGGCGTGTACACAGAGGCTCCCTGTAGTTCGGGACTGGATCAAG 2613
```

```
2528 CTCAGAGGACCAAGCCAGCGGTGTACACAGGCTCCCTCTGTTTCGGGACTGGATCAAG 2587
2614 AGCAGACTGGGGTATAGCAGCATGACAGACA 2645
2588 AGAACACTGGGGTATAGGGGCCCGGGGACCCA 2619

RESULT 9
AB030036 3128 bp mRNA linear PRI 06-JAN-2001
LOCUS Homo sapiens mRNA for prostamin, complete cds.
DEFINITION AB030036
ACCESSION AB030036
VERSION 1 GI:12249014
KEYWORDS prostamin.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (sites)
AUTHORS Yamaguchi, N. and Mitsui, S.
TITLE Molecular cloning of a novel transmembrane serine protease
expressed in human prostate
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 3128)
AUTHORS Yamaguchi, N. and Mitsui, S.
TITLE Direct Submission
JOURNAL Submitted (14-JUL-1999) Nozomi Yamaguchi, Kyoto Prefectural
University of Medicine, Res. Ins. Geriatrics; Kawarachi Hirokoji,
Kyoto, Kyoto 602-8566, Japan (E-mail:nozomi@koto.kpu-m.ac.jp,
Tel:81-75-251-5848, Fax:81-75-251-5848)
FEATURES
Location/Qualifiers
1..3128
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/tissue_type="prostate"
16..2583
/codon_start=1
/product="prostamin"
/protein_id="BAB20376.1"
/db_xref="GI:12249015"
/translation="MGSDRARKGGGPKDFGAGLKYNSRHEKVNGLGEVFEFLPVNVV
KXVEHGRVWVLAALVILGILGFLVHLQYRDVVRQKVFNGYMRITNFV
IYNHLSSTEFVSLASKVDALKLYSGVPLGPKHESAVTAFSEGVIAVYWSFES
IQHLEVEARVMAERVVMPPARSLKSFVTSVVAPTQSKVQRPQDNCSRGL
HARGVELRPTTPGPDSPYPAHRCQWALRGDADSVLSLTFSEDLASCDERGSDLV
TVNTHLSMEPHALVOLCGTYPPSNLTHSSHNSONLITLITNTERHGFPEATFPOL
PMSSCGGRLKAQGTFSPPYHPNIDCTNIEVNNQHVKVRFFKFFLLBEGV
PAGTCPKDYVEINGEKYCGERSQFVTSNSKITVRFHSDQSYDTDTGLAEYLSYDS
DPCQGTCTGTCIRKELRCGDMADCTHSDLCNCSGADAGHFTCKNFKPLFWVC
DSVNDGDNSSDEQSCPAOTFCGNGKCLSKSQCNKDCGDCGDSDEASCPLKVVVT
CTKHVRCNLGLCLSKGNPCDGEDSGDEKDCDGLRSTQARVVGGTADDEG
EWPQVSLHALQGHICASLISPNWLYSAHCYIDDRGFYSDDTQWTVFLGHDQS
QKSAFGVQRRLRIISHPFNDFTQDILALLELEKPAEYSSMRPCLIPDASHVFP
AKALWVTGWHTQYGTGALIQRELVRINQTTCEMLLPQIITPRMVCVGLSGV
DSCQSDSGPLSSVEADGRIFQAGVVGWDGCAQRNKPQVYTRLPFRDWIKENTGV"
BASE COUNT 637 a 954 c 956 g 581 t
ORIGIN

Query Match 60.7%; Score 1884.6; DB 9; Length 3128;
Best Local Similarity 81.4%; Pred. No. 0;
Matches 2222; Conservative 0; Mismatches 504; Indels 5; Gaps 3;

QY 50 GACCGCCAAAACCATGGGTAGCAATCGGGCGCGCAAGCGCGGAGGGGCTCTCAGACTT 109
Db 3 GGCCTCGGGGACATGGGAGGATCGGGCCCGCAAGGGCGAGGGCGGCCAAGACTT 62
QY 110 CGGCGCGGACTCAAGTACAATCCCGGCTAGAACATGAATGGCTTTGAGAGGCTT 169
Db 63 CGGCGCGGACTCAAGTACAATCCCGGCTAGAACATGAATGGCTTTGAGAGGCTT 122
QY 170 GGAGTTCCTGCTCGAACAATGCCAAGAAAGTGAGAGAGCGAGGCCCGCGCTGGGT 229
```


QY 2390 TGAGACCTCATGCGCAGCAGATCACCCACGAATGATGTGTGGGTTTCCTCAGTGG 2449
| | | | |
Db 2343 CGAGAACCTCTCGCGAGCAGATCAACCCCGCATGATGTGGGCTTCCTCAGCGG 2402
| | | | |
QY 2450 GGGTGTGAGCTCTCGCAGGGTGACTCTGTGGGCCCTTGTCAAGCGCGGAGAAAGATGG 2509
| | | | |
Db 2403 CGGCGTGACTCTCTCGCAGGGTGATTCGGGGGACCCCTGTCCAGCGTGGAGGCGGATGG 2462
| | | | |
QY 2510 GCGAATGTTCCAGAGCTGGTGTGGTGAAGCTGGGGTGAAGCTGGCCCTCAGAGGAACAGCC 2569
| | | | |
Db 2463 GCGGATCTTCAGGCGCGGTGTGGTGAAGCTGGGGGAGACGGCTGGCCCTCAGAGGAACAGCC 2522
| | | | |
QY 2570 AGGCGGTATACAAAGGCTCCCTGTAGTTGGGCACTGGATCAAGAGCAGACACTGGGGTATA 2629
| | | | |
Db 2523 AGGCGGTATACAAAGGCTCCCTGTGGTGAAGCTGGATCAAGAGGAACACTGGGGTATA 2582
| | | | |
QY 2630 GCAGCATGGAAGCAGCAGCGGACCAACAAACCCACAGGATGCGCGCATGCACACCTGG 2689
| | | | |
Db 2583 GGGCGCGGG---GCCACCCAAATGTATACCTGGGGGCCACCCATCGTCCACCCAGT 2639
| | | | |
QY 2690 ATACAGGAGGACACTGACGACATTTATGCTGTGGCTTCCCTCCCTCCACACCAACCC 2749
| | | | |
Db 2640 GTGCAGC-CCTGCGAGCTGGAGACTGGACCGCTGACTGCACACGCGCC-CCAGAACATA 2697
| | | | |
QY 2750 GACTGTGAATGCATCTCTTAGGACTCAGAGT 2780
| | | | |
Db 2698 CACTGTGAATCAATCTCCAGGGCTCCAAAT 2728
| | | | |
RESULT 10
AR081724
LOCUS AR081724 3147 bp DNA linear PAT 31-AUG-2000
DEFINITION Sequence 1 from patent US 5972616.
ACCESSION AR081724
VERSION AR081724.1 GI:10008450
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE
1 (bases 1 to 3147)
O'Brien, T.J. and Tanimoto, H.
TAGD-15: an extracellular serine protease overexpressed in breast
and ovarian carcinomas
JOURNAL Patent: US 5972616-A 1 26-OCT-1999;
FEATURES Location/Qualifiers
source 1..3147
BASE COUNT 654 a 952 c 958 g 583 t
ORIGIN
Query Match 60.6%; Score 1883.2; DB 6; Length 3147;
Best Local Similarity 81.2%; Pred. No. 0;
Matches 2223; Conservative 0; Mismatches 508; Indels 5; Gaps 3;
QY 45 GATCGACCGCCAAAACCATGGGTAGCAATCGGGCCCGCAAGCGCGGAGGGGCTCTCAG 104
| | | | |
Db 5 GAGCGGCTCGGGGTACCATGGGAGCGATCGGGCCCGCAAGGCGGAGGGGCGCCGAAG 64
| | | | |
QY 105 GACTTCGGCGGGGACTCAAGTACAACTCCCGGTAGACACATGAATGGCTTTGAGGAG 164
| | | | |
Db 65 GACTTCGGCGGGGACTCAAGTACAACTCCCGGTAGACACATGAATGGCTTTGAGGAG 124
| | | | |
QY 165 GGTGTGAGTTCCTGCTCGCAACAACTGCCAAGAAAGTGGAGAGCGAGGCCCGCAGCGC 224
| | | | |
Db 125 GCGTGTGAGTTCCTGCTCGCAACAACTGCCAAGAAAGTGGAGAGCGATGGCCGGGCGC 184
| | | | |
QY 225 TGGTGTGTGTGGGAGTGTCTGTTCAGTTCCTCTGCTCTCCCTCATGGCTGGCTTG 284
| | | | |
Db 185 TGGTGTGTGTGGGAGCGGTCTGATCGGCTCTCTGCTGCTGCTGGGATCGGCTTC 244
| | | | |
QY 285 CTGCTGTGGCACTTCCATTATCGAATGTGGGGTTCAAAAGTCTTCATGGCCATCTG 344
| | | | |

Db 245 CTGCTGTGGCAATTTGCAGTACCGGGACGTGCTCCAGAGGCTCTTCAATGGCTACATG 304
QY 345 AGGATCAAAATGAGATCTTTCTGATGCGTATGAGAACTCCACCTCCACAGAGTTTATC 404
| | | | |
Db 305 AGGATCAAAATGAGATCTTTCTGATGCGCTTACGAGAACTCCAACTCACTGAGTTGTA 364
| | | | |
QY 405 AGCTTGCCAGCAGCAGGTGAAGGAGGCGCTCAAGCTGCTGACAAATGAAGTCCCTGCTCTG 464
| | | | |
Db 365 AGCTTGCCAGCAGCAGGTGAAGGAGGCGCTCAAGCTGCTGACGAGGATGCCATTCCTG 424
| | | | |
QY 465 GGTCTCTTACCACAAGAGTCTGCTTAACCTTCACTGAGGCGAGTGTCACTGCGCTAC 524
| | | | |
Db 425 GGGCCCTTACCACAAGAGTCTGCTGACGCGCTTACGCGAGGCGAGCGTCACTGCGCTAC 484
| | | | |
QY 525 TACTGTGAGAGTTCAGCATCCCGCCACACACTGCGCAGAGAGGTTGATCGGCGCATGGCT 584
| | | | |
Db 485 TACTGTGCTGAGTTCAGCATCCCGCCACACACTGCGTGGAGGAGGCGGCGCTCATGGCC 544
| | | | |
QY 585 GTGAGCGAGTTCATCAATTCGCCACCCCGAGCAGCGGCACTGAAATCCTTCTGCTAAACA 644
| | | | |
Db 545 GAGGAGCGGCTAGTCTATGCTGCGCGCGCGCGCTCCCTGAAGTCTTCTGTTGGTCAAC 604
| | | | |
QY 645 TCTGTGTGCTGCTTCCCGATTTGACCCCAAGAAATGCTGCGAGGAGTCTCAGGACAAAGCTGC 704
| | | | |
Db 605 TCACTGTGTGCTTTCGCCACCGACTCCAAAACAGTACAGAGGACCCAGGAGCAACAGCTGC 664
| | | | |
QY 705 AGTTTTCCTGCTGATGCGCATGCGGATGGTGCAGAGTGCACAGCTTCACTACCCCTGCTTCCC 764
| | | | |
Db 665 AGCTTGTGCGCTGACGCGCGCGGTGTGGAGCTGATGCGTTCACACGCGCGGCTTCCCT 724
| | | | |
QY 765 AACAGTCCCTTACCCGCGCATGCGCTGCTGCTGCTGCTGCGGGGAGCGCGACTCT 824
| | | | |
Db 725 GACAGCCCTTACCCGCTCATGCGCGCTGCTGCTGCTGCGGGGAGCGCGACTCA 784
| | | | |
QY 825 GTCTGAGCTTCACTTCCGAACTTGTGATGCTGCTGCTGCTGCTGATGAGCATGCGAGTGC 884
| | | | |
Db 785 GTCTGAGCTTCACTTCCGAGCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 844
| | | | |
QY 885 CTGCTGACCGTGTATGATGCTGAGCGCTGAGCGCCATGGAACCCCGCTGCTGCTGCTGCTGCTG 944
| | | | |
Db 845 CTGCTGAGCGGTGACAAACCCCTGAGCGCCATGAGCGCCCGCTGCTGCTGCTGCTGCTGCTG 904
| | | | |
QY 945 GGCACCTTCTCACCCCTCTACAACTGACTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTG 1004
| | | | |
Db 905 GGCACCTTACCT 964
| | | | |
QY 1005 ACCTGATTAACCAATGCTGACCGCGAGCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1064
| | | | |
Db 965 ACATGATTAACCAACATGAGCGCGCGCATCTCCCGCTTTTGGAGCCACCTCTCTCTCTCTG 1024
| | | | |
QY 1065 CCCAAGATGAGCAGCTGTGGCGCTTTTGGTGAACACCCCAAGGACATTTAGCAGCC 1124
| | | | |
Db 1025 CTTAGATGAGCAGCTGTGGAGCCGCTTACGTAAAGCCGAGGAGACATTTCAACAGCC 1084
| | | | |
QY 1125 TACTATCCAGGCGCACTACCGCCCAACATCACTGACATGCAATGGAATATCAAGGTGCCCAAC 1184
| | | | |
Db 1085 TACTATCCAGGCGCACTACCGCCCAACATGACTGACATGGAATATGAGGTGCCCAAC 1144
| | | | |
QY 1185 AACGGGAGCTGAGGTGCGCTTCAAACTTCTATCTGCTGAGCCCAACAGTACCAAGT 1244
| | | | |
Db 1145 AACCGATGTGAGGTGAGCTTCAAACTTCTATCTGCTGAGCCCGCGCTGCTGCTGCTGCTG 1204
| | | | |
QY 1245 GGTCTCTGACCAAGGACTATGTGGAGATCAACCGGAGAAAGTACTGCGGTGAGAGGTCC 1304
| | | | |
Db 1205 GGCACCTGCCCCAAGGACTACGTGGAGATCAATGGGAGAAATATCTGCGGAGAGAGTCC 1264
| | | | |
QY 1305 CAGTTTGTGTGAGCAGCAACAGCAGCAAGATTAAGTCCACTTCCATTTCTGATCACTG 1364
| | | | |
Db 1265 CAGTTTGTGTGAGCAGCAACAGCAGCAAGATTAAGTCCACTTCCATTTCTGATCACTG 1324
| | | | |
QY 1365 TACACGAGCAGCGGCTTCTGAGTGTGAGTCTCTCTCTGAGTCCAGCAGCCGCTGCCA 1424
| | | | |
Db 1325 TACACGAGCAGCGGCTTCTGAGTGTGAGTCTCTCTCTGAGTCCAGTCCAGTCCAGTCCAG 1384
| | | | |

QY 1425 GGGATGTTTCATGTCAAGACTGAGCGGTGCAATCCGAAAGGAACCTGCGCTGCGACGGCTGG 1484
DB 1385 GGGCAGTTTCAGTCCGCGACCGGGGGGTGTATCCGGAAGAGCTGCGCTGTGATGGCTGG 1444
QY 1485 GCAGACTGCCCGGATATATAGTGTAGAGGCTTACTGCCGATGCAATGCCACCCACAGATTC 1544
DB 1445 GCCGACTGCACCGACACAGCGATGAGTCAACTGCAGTTCGACGCGCGGCCACAGATTC 1504
QY 1545 ACGTGCAAAAACCAAGTTCCTGCAAGCCCTCTTCTGGGTCTGTGACAGTGTCAAGCACTGT 1604
DB 1505 ACGTGCAAGAACAGTTCCTGCAAGCCCTCTTCTGGGTCTGTGACAGTGTGAAACACTGTC 1564
QY 1605 GGGAGCGGAAGTGTGACGAGAGGGCTGTGACGTGTCTCTGCTGGGAGTTTCAAGTGTTCGAAT 1664
DB 1565 GGAGACAACAGCGACGAGCGAGCGGTGCAGTGTGTCGGGCCAGACCTTCAGGTGTTCGAAT 1624
QY 1665 GGGAGTGTCTCCCTCAGAGCCAGAGTGTATGGAAGGACAACTGTGAGATGGGTCT 1724
DB 1625 GGGAGTGCCTCTCGAAAGCCAGCAGTGCATGCAATGGGAAGGACGACTGTGGGAGCGGTCC 1684
QY 1725 GACGAGGCTTCATGTGACAGCGTGAATGTCTCTTTCACCAAAATATACCTACCGCTGC 1784
DB 1685 GACGAGGCTCTCTGCCCAAGGTGAACTGTCTGCTTGTACCAACACACCTTACCGCTGC 1744
QY 1785 CAAATAGGCTCTGTCTGAGCAAGGCAACCTGAGTGTGATGGGAAGACGAGCTGTAGC 1844
DB 1745 CTCAATGGCTCTGCTTGAGCAAGGCAACCTGAGTGTGACGGAAGGAGGACTGTAGC 1804
QY 1845 GATGGCTCCGATGAGAAACCTGTGACTGTGGCTGCGATCTCTTACCAACACAGCTGC 1904
DB 1805 GACGGCTCAGATGAGAGGACTGTGACTGTGGCTGCGGTCAITTCAGAGACAGGCTGT 1864
QY 1905 GTGTTGTGTGACCAATGCGGACGAGGCGAGTGGCTCGCAGGTGAGCTCCACAGCC 1964
DB 1865 GTTGTGGGGCAAGATGCGATGAGGCGAGTGGCTCGCAGGTGAGCTCCAGTCTGCT 1924
QY 1965 CTGGGCGAGGGCCACTGTGTGGGGCTCTGCTCACTCTCTGCTGCTGCTGCTGCTGCA 2024
DB 1925 CTGGGCGAGGGCCACTGTGGGTGCTTCCCTCATCTCTCCCACTGCTGCTGCTGCTG 1984
QY 2025 GCTCATGCTTTTCAGATGACAAAATTTCAAGTACTCAGACTACAGATGTGACGGCC 2084
DB 1985 GCACACTGCTATCATGATGACAGAGGATTCAGGTACTCAGACCCACCGCAGTGGACGGCC 2044
QY 2085 TTCTGTGGTCTGTGGACGACGAGCAAGCGAGTGCCTCTGGGGTGCAGGAGCTCAAGCTC 2144
DB 2045 TTCTGTGGCTGTGACGACGACGAGCCAGCGAGCGCCCTGGGGTGCAGGAGCGCAGCTC 2104
QY 2145 AACGATATCATCACCCCTTCTTCAATGATTTCACTTCGACTATGACATCGCCTTG 2204
DB 2105 AAGCGCATCATCTCCCACTTCTTCAATGACTTCACTTCGACTATGACATCGCGCTG 2164
QY 2205 CTGAGCTGGGAAGTGTGGAGTACAGACCGTGTGCGCCCATCTGCTGCTGCTGCTGAT 2264
DB 2165 CTGAGCTGGGAAGAACCGGAGTACAGCTTCATGTTGGGCGCCCATCTGCTGCTGCTG 2224
QY 2265 GCTACCATGCTTTCTCTGTGGCAAGGCCACTGTGGGTCAAGGCTGGGGGACACAAAA 2324
DB 2225 GCCTCCCATGCTTCTCCCTGCGGCAAGGCCATCTGGGTCAAGGCTGGGACACACCCAG 2284
QY 2325 GAGGAGGTACCGGAGGCTGATCTCTGAGAGGGTGTGATCTCTGTGTATCAACAGACCC 2384
DB 2285 TATGGAGGCACTGGCGGCTGATCTCTGCAAAAGGTTGAGATCCCGCTCATCAACAGACC 2344
QY 2385 ACCTGTGAGGACCTCATGCCAGCAGATCAACCCCAAGATGATGTGTGGTTCCTC 2444
DB 2345 ACCTGTGAGGACCTCTGCGCAGCAGATCAACCCCGGCTGATGTGTGGTTCCTC 2404
QY 2445 AGTGGGGTGTGGACTCTCTGCGAGGGTGAATCTGTGCGCCCTTGTCAAGCGCGGAGAA 2504
DB 2405 AGCGGCGGTGGACTCTGTCAGGGTGAATCTGCGGGGACCCCTGTTCAGCGGTGGAGCG 2464

QY 2505 GATGGCGAAATGTTCCAGGCTGGTGTGTGAGCTGGGTGAAGCTGCGCTCAGAGGAAC 2564
DB 2465 GATGGCGGATCTTCAGGCGGTGTGTGAGCTGGGAGACGCTGCGCTCAGAGGAAC 2524
QY 2565 AAGCCAGCGGTGTACAAAGGCTCCCTGTGTGTTCCGGGACTGTGATCAAGAGCACTGGG 2624
DB 2525 AAGCCAGCGGTGTACAAAGGCTCCCTGTGTGTTCCGGGACTGTGATCAAGAGCACTGGG 2584
QY 2625 GTATAGCAGCATGTGACAGACGCGACCAAAACCCACAGGATGCCGACATGCA 2684
DB 2585 GTATAGGCGCGGG---GCCACCCAAATGTGTACCTTGGGGGCCACCCATCTGCTCCACC 2641
QY 2685 CCTGGATACAGGAGAGGAACACTGACGACATTTATGTGTGGCTTCCCTCCCAACACA 2744
DB 2642 CCAGTGTGCAG-CCTGACGCTGGAGCTGGACCGCTGACTGACACCGGCCCC-CCAGA 2699
QY 2745 ACCCAGACTGTGAACCTGATCTTAGACTCAGAGT 2780
DB 2700 ACATACACTGTGAACCTCAATCTCCAGGGCTCCAAAT 2735

RESULT 11
AR229704 Locus AR229704 3147 bp DNA linear PAT 20-DEC-2002
DEFINITION Sequence 1 from patent US 6451500.
ACCESSION AR229704
VERSION AR229704.1 GI:27269434
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 3147)
AUTHORS Leon,J.W.
TITLE Imaging member containing heat switchable carboxylate polymer and method of use
JOURNAL Patent: US 6451500-A 1 17-SEP-2002;
FEATURES Location/Qualifiers
source 1. 3147
BASE COUNT 654 a 952 c 958 g 583 t
ORIGIN

Query Match 60.6%; Score 1883.2; DB 6; Length 3147;
Best Local Similarity 81.2%; Pred. No. 0;
Matches 2223; Conservative 0; Mismatches 508; Indels 5; Gaps 3;

QY 45 GATCGGACCCCAAAACCATGGGTAGCAATCGGGCCGCGAAGCCGAGGGGGTCTCAG 104
DB 5 GAGCGGCTCGGGGTACCATGGGAGCGATCGGGCCCGCAAGGGCGAGGGGGCCCAAG 64
QY 105 GACTTCGGCGCGGACTCAAGTACAACTCCCGCTAGAGAACATGAATGGCTTGGAGAG 164
DB 65 GACTTCGGCGCGGACTCAAGTACAACTCCCGCTAGAGAACATGAATGGCTTGGAGAG 124
QY 165 GGTGTGAGTTCCTGCTGCAACATGCGCAAGAAAGTGGAGAGCGAGGCCCGCGCC 224
DB 125 GGGCTGAGTTCCTGCTGCAACATGCGCAAGAAAGTGGAGAGCGAGGCCCGCGCC 184
QY 225 TGGTGTGTGCTGGTGGCAGTGTGTTGAGTCTCTCTGCTCTCTCTCTCTCTCTCTCT 284
DB 185 TGGTGTGTGCTGGCAGCGGTGATCGGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 244
QY 285 CTGCTGTGGCACTTCCATATCGGAATGTGCGGGTTCAAAAAGTCTTCAATGGCACTG 344
DB 245 CTGCTGTGGCACTTCCATATCGGAATGTGCGGGTTCAAAAAGTCTTCAATGGCACTG 304
QY 345 AGGATCAAAATCAGATCTTCTGAGTGTGATGAGAACTCCACCTCCACAGAGTTTATC 404
DB 305 AGGATCAAAATCAGATCTTCTGAGTGTGATGAGAACTCCACCTCCACAGAGTTTATC 364
QY 405 AGCTGGCGGACGAGTGAAGAGGGCGCTGAACTGCTGTACATGAGTCCCTGCTGCTG 464
DB 365 AGCTGGCGGACGAGTGAAGAGGGCGCTGAACTGCTGTACATGAGTCCCTGCTGCTG 424

QY 465 GGTCCCTTACCACAAGAGTGGCTGTAACTGCTTCACTGAGGCGAGTGTATCGCCCTAC 524
DB 425 GGCCCTTACCACAAGAGTGGCTGTAACTGCTTCACTGAGGCGAGTGTATCGCCCTAC 484
QY 525 TACTGTGACAGTTTACAGTCCCTCCCAACACTGCGCAGAGAGTGTATCGCCCTAC 584
DB 485 TACTGTGACAGTTTACAGTCCCTCCCAACACTGCGCAGAGAGTGTATCGCCCTAC 544
QY 585 GTGAGAGAGTGTAACTTGGCACCACCGGAGCAGGCGACTGAACTCCTTGTGCTTAA 644
DB 545 GAGAGAGAGTGTAACTTGGCACCACCGGAGCAGGCGACTGAACTCCTTGTGCTTAA 604
QY 645 TCTGTGTGCGCTTCCCTTACAGTCCCTCCCAACACTGCGCAGAGAGTGTATCGCCCTAC 704
DB 605 TCACTGTGCGCTTCCCTTACAGTCCCTCCCAACACTGCGCAGAGAGTGTATCGCCCTAC 664
QY 705 AGTTTTCCTTGCATGCGCATGTTGAGCAGTGTACAGCTTCACTACCTTGGCTTCCC 764
DB 665 AGCTTTGGCTGACAGCGCGGCTGTGAGCTGTATGCGCTTCACTACCGCGGCTTCCC 724
QY 765 AACAGTCCCTTACCGGCGCATGCGCGTGTGAGTGTGAGCTGTATGCGCTTCACTACCG 824
DB 725 GACAGCCCTTACCGGCGCATGCGCGTGTGAGTGTGAGCTGTATGCGCTTCACTACCG 784
QY 825 GTGCTGAGCTTCACTTCCGAGCTTGTGAGTGTGAGCTGTATGCGCTTCACTACCG 884
DB 785 GTGCTGAGCTTCACTTCCGAGCTTGTGAGTGTGAGCTGTATGCGCTTCACTACCG 844
QY 885 CTGCTGACCGTGTATGATAGCTTACCGGCGCATGCGCGTGTGAGCTGTATGCGCTT 944
DB 845 CTGCTGACCGTGTATGATAGCTTACCGGCGCATGCGCGTGTGAGCTGTATGCGCTT 904
QY 945 GGCACCTTCTCACCTTCCCTTACAGTGTGAGTGTGAGCTGTATGCGCTTCACTAC 1004
DB 905 GGCACCTTACCTTCCCTTACAGTGTGAGTGTGAGCTGTATGCGCTTCACTAC 964
QY 1005 ACCTGTATTAACCTTACAGTGTGAGTGTGAGCTGTATGCGCTTCACTAC 1064
DB 965 ACCTGTATTAACCTTACAGTGTGAGTGTGAGCTGTATGCGCTTCACTAC 1024
QY 1065 CCAAGATGAGAGCTGTGCGGCTTTTGTGAGTGTGAGCTGTATGCGCTTCACTAC 1124
DB 1025 CCAAGATGAGAGCTGTGCGGCTTTTGTGAGTGTGAGCTGTATGCGCTTCACTAC 1084
QY 1125 TACTATCCAGGCGCTTACCGGCGCATGCGCGTGTGAGTGTGAGCTGTATGCGCTT 1184
DB 1085 TACTATCCAGGCGCTTACCGGCGCATGCGCGTGTGAGTGTGAGCTGTATGCGCTT 1144
QY 1185 AACGGAGCTGAGTGTGCGGCTTTTGTGAGTGTGAGCTGTATGCGCTTCACTAC 1244
DB 1145 AACGGAGCTGAGTGTGCGGCTTTTGTGAGTGTGAGCTGTATGCGCTTCACTAC 1204
QY 1245 GGCTTCTGACAGGAGTGTGAGTGTGAGTGTGAGCTGTATGCGCTTCACTAC 1304
DB 1205 GGCTTCTGACAGGAGTGTGAGTGTGAGTGTGAGCTGTATGCGCTTCACTAC 1264
QY 1305 CAGTTTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGCTGTATGCGCTTCACTAC 1364
DB 1265 CAGTTTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGCTGTATGCGCTTCACTAC 1324
QY 1365 TACAGGAGACCGGCTTCTTGTGAGTGTGAGTGTGAGCTGTATGCGCTTCACTAC 1424
DB 1325 TACAGGAGACCGGCTTCTTGTGAGTGTGAGTGTGAGCTGTATGCGCTTCACTAC 1384
QY 1425 GGGATGTTTCAAGTGTGAGTGTGAGTGTGAGTGTGAGCTGTATGCGCTTCACTAC 1484
DB 1385 GGGATGTTTCAAGTGTGAGTGTGAGTGTGAGTGTGAGCTGTATGCGCTTCACTAC 1444
QY 1485 GCAGACTGCCCGGATTTAGTGTGAGTGTGAGTGTGAGCTGTATGCGCTTCACTAC 1544
DB 1445 GCGACTGTCACCGACACAGAGTGTGAGTGTGAGTGTGAGCTGTATGCGCTTCACTAC 1504

QY 1545 ACSTGCAAAAAACAGTCTTCTGCAAGCCCTTCTTCTGGGTCTGTGACAGTGTCAACAGCTGT 1604
DB 1505 ACSTGCAAAAAACAGTCTTCTGCAAGCCCTTCTTCTGGGTCTGTGACAGTGTCAACAGCTGT 1564
QY 1605 GGGGACGGAAGTGTGACAGGAGGCTGAGTGTCTTCTGGGTCTGTGAGTGTCAAGTGTCAAT 1664
DB 1565 GGGGACGGAAGTGTGACAGGAGGCTGAGTGTCTTCTGGGTCTGTGAGTGTCAAGTGTCAAT 1624
QY 1665 GGGGAGTGTCTTCTCTCAGAGCCAGAGTGTAAAGGAGGAGCAACTGTGAGATGGGTCT 1724
DB 1625 GGGGAGTGTCTTCTCTCAGAGCCAGAGTGTAAAGGAGGAGCAACTGTGAGATGGGTCT 1684
QY 1725 GACAGGCTTCACTGTGACAGGCTGAACTGTCTTCTTGCACCAAAATATACCTACCGCTGC 1784
DB 1685 GACAGGCTTCACTGTGACAGGCTGAACTGTCTTGTACCAAAATATACCTACCGCTGC 1744
QY 1785 CAAAATGGGCTTCTGTGACAGGCTGAACTGTCTTGTACCAAAATATACCTACCGCTGC 1844
DB 1745 CTAATGGGCTTCTGTGACAGGCTGAACTGTCTTGTACCAAAATATACCTACCGCTGC 1804
QY 1845 GATGGCTTCCGATGAGAAAAAATGTGACTGTGGCTGCGATCTTCTTACCAAAACAGGCTGC 1904
DB 1805 GACGCTTCACTGTGAGAGGAGCTGCGACTGTGAGCTGCGGTCTTCTACGAGACAGGCTGC 1864
QY 1905 GTGGTGTGTGACAGAAATGCGGACGAGGCGAGTGTGCGCTTGTGAGTGTGAGCTTCAAGCC 1964
DB 1865 GTGGTGTGTGACAGAAATGCGGACGAGGCGAGTGTGCGCTTGTGAGTGTGAGCTTCAAGCC 1924
QY 1965 CTGGGCGAGGCGCTTGTGTGGGCTGCTCATCTCTCTGAGTGTGAGTGTGAGCTTCAAGCC 2024
DB 1925 CTGGGCGAGGCGCTTGTGTGGGCTGCTCATCTCTCTGAGTGTGAGTGTGAGCTTCAAGCC 1984
QY 2025 GCTCATGTCTTCTCAGGATGACAAAAATTTCAAGTACTCAGACTTACAGTGTGAGGCGCC 2084
DB 1985 GCACTGTCTACATCTGATGACAGAGATTTCAAGTACTCAGACTTACAGTGTGAGGCGCC 2044
QY 2085 TTCTGTGGTCTGTGACAGAGAGGCGAGTGTCTTGTGGGTGTGAGGCTGTGAGGCTGT 2144
DB 2045 TTCTGTGGTCTGTGACAGAGAGGCGAGTGTCTTGTGGGTGTGAGGCTGTGAGGCTGT 2104
QY 2145 AAAGTGTATCATACCAACCTTCTTCAATGATTTCACTTCACTTCACTTCACTTCACTTCACT 2204
DB 2105 AAGCGATCATCTTCCCACTTCTTCAATGATTTCACTTCACTTCACTTCACTTCACTTCACT 2164
QY 2205 CTGGAGCTGTGAGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGT 2264
DB 2165 CTGGAGCTGTGAGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGT 2224
QY 2265 GCTACCTATGTCTTCTGTGAGAGGCGCTTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGT 2324
DB 2225 GCTACCTATGTCTTCTGTGAGAGGCGCTTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGT 2284
QY 2325 GAGGAGGTGTGAGAGGCTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGT 2384
DB 2285 TATGAGGAGCTGTGAGAGGCTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGT 2344
QY 2385 ACCTGTGAGAGCTGTGAGAGGCTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGT 2444
DB 2345 ACCTGTGAGAGCTGTGAGAGGCTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGT 2404
QY 2445 AGTGGGGGTGTGAGAGGCTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGT 2504
DB 2405 AGTGGGGGTGTGAGAGGCTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGT 2464
QY 2505 GATGGGCGAATGTGAGAGGCTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGT 2564
DB 2465 GATGGGCGAATGTGAGAGGCTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGT 2524
QY 2565 AAGCCAGGCGTGTACAAAGGCTTCTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGT 2624
DB 2525 AAGCCAGGCGTGTACAAAGGCTTCTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGT 2584
QY 2625 GTATAGCAGTGTGAGAGGCTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGT 2684

Db 2585 GTATAGGGCCGGG--GCCACCAAAATGTGTACACCTGCGGGGCCACCCATCTCTCCACC 2641
Qy 2585 CCTGGATACAGGAGAGAACTGACAGACATTTATGTGTGGCTCTCCCGCCCAACACA 2744
Db 2642 CCAAGTGTGACG-CCTGAGGCTGGAGACTGGACCGCTGACTGACCAAGCGCC-CCAGA 2699
Qy 2745 ACCCAGACTGTGAATGATGATCTTAGACTCAGAGT 2780
Db 2700 ACATACACTGTGAATCAATCTCAGGGCTCAAAAT 2735

RESULT 12
LOCUS AR229712/c 3147 bp RNA linear PAT 20-DEC-2002
DEFINITION Sequence 18 from patent US 6451500.
ACCESSION AR229712
VERSION AR229712.1 GI:27269442
KEYWORDS
SOURCE
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 3147)
AUTHORS Leon, J.W.
TITLE Imaging member containing heat switchable carboxylate polymer and method of use
JOURNAL Patent: US 6451500-A 18 17-SEP-2002;
FEATURES Location/Qualifiers
source 1..3147
BASE COUNT 583 a 958 c 952 g 654 t
ORIGIN

Query Match 60.6%; Score 1883.2; DB 6; Length 3147;
Best Local Similarity 81.2%; Pred. No. 0;
Matches 223; Conservative 0; Mismatches 508; Indels 5; Gaps 3;

Qy 45 GATCGGACCGCAAAACATGGGTAGCAATCGGGCGCGCAAGCGCGGAGGGGGCTCTCAG 104
Db 3143 GAGCGGCTCGGGGTACCATGGGAGCGATCGGCCCGCAAGGCGGAGGGGCCGAG 3084
Qy 105 GACTTCGGCGGGGACTCAAGTAACTCCGGGTAGAGAACATGAATGGCTTTGAGGAG 164
Db 3083 GACTTCGGCGGGGACTCAAGTAACTCCGGGTAGAGAACATGAATGGCTTTGAGGAA 3024
Qy 165 GGTGTGGAGTTCTCGCTCGCAAAATGCCAAGAAAGTGGAGAGCGAGCCCGCAGCGC 224
Db 3023 GGCCTGGAGTTCTCGCAGTCAACAGCTCAAGAAAGTGGAAAGCATGGCCCGGGCGC 2964
Qy 225 TGGGTGTGTGTGGGAGTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGT 284
Db 2963 TGGGTGTGTGTGGGAGCGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2904
Qy 285 CTGTGTGGCACTTCATATCGGAATGTGGGGTTCAAAAGTCTTCAATGGCCATCTG 344
Db 2903 CTGTGTGGCACTTCATATCGGAATGTGGGGTTCCAGAGGTCTTCAATGGGTACATG 2844
Qy 345 AGGATCAAAATGAGATCTTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 404
Db 2843 AGGATCAAAATGAGATCTTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2784
Qy 405 AGCCTGGCAGCCAGGTGAAGGAGCGCTGAAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 464
Db 2783 AGCCTGGCAGCAAGGTGAAGGAGCGCTGAAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2724
Qy 465 GGTCTTACCAAGAGTGTGGTGTAACTGCTTTCAGTGTGGGAGGAGTGTATCGGCTAC 524
Db 2723 GGCCCTTACCAAGAGTGTGGTGTAACTGCTTTCAGTGTGGGAGGAGTGTATCGGCTAC 2664
Qy 525 TACTGGTTCAGAGTTTCAGATCCCGCCACACCTGGCAGAGAGGTGTATCGGCTACGT 584
Db 2663 TACTGGTTCAGATCCCGCCACACCTGGTGTGGGAGGAGGCGGCTCATGGCC 2604

Qy 585 GTGAGCGAGTTGTAACTTTGCCACCCCGGACCGGACATGAAATCTTCTGCTGCTAACA 644
Db 2603 GAGAGCGCGTAGTGTATGCTTGCCTCCCGGGCGCGCTCCCTGAAAGTCTTGTGTGTCACC 2544
Qy 645 TCTGTGTGTGCTTCTCCCATTTGACCCCGAGAAATGTGTGAGAGGACTCAGGACCAACAGTGC 704
Db 2543 TCAGTGTGTGCTTCTCCCGACCGACTCCAAAACAGTACAGAGGACCCAGGACCAACAGTGC 2484
Qy 705 AGTTTGTGCTGTATGCTGATGTGTGAGAGTGTGACAGCTTCTACTACCTGCTGGCTTCCCC 764
Db 2483 AGCTTTGGCCTGCACGCCCGCGGTGTGTGAGCTGTATGCGCTTACACAGCCCGGGTTCCT 2424
Qy 765 AACAGTCCCTACCGCGGCATGCCCGCTGCAGTGGGTCTCTGCGGGGGAGCGGACTCT 824
Db 2423 GACAGCCCTTACCCCGCTGATGCCCGTGCAGTGGGCCCTTGGGGGGAGCCCGACTCA 2364
Qy 825 GTGCTGAGCCTCACCTTCCGAAGCTTTGATGTGTGCTCCCTGTGTATGAGCATGGCAGTGAC 884
Db 2363 GTGCTGAGCCTCACCTTCCGAGCTTTGACCTTGTGCTCTGCGACGAGCGCGCAGCGAC 2304
Qy 885 CTGGTCACTGCTGTATGATAGCTGAGCCCATGGAACCCACAGCTGTGTGTGCGGTGTGT 944
Db 2303 CTGGTCACTGCTGTATGATAGCTGAGCCCATGGAACCCACAGCTGTGTGTGCGGTGTGT 2244
Qy 945 GGCACCTTCTCACCTCTCTACACCTGACTTTCTCTCTCTCCCGAGAACGCTTCTCTGTGTC 1004
Db 2243 GGCACCTTCTCTCTCTACACCTGACTTCTCTCTCTCCCGAGAACGCTTCTCTGTGTC 2184
Qy 1005 AGCTGTATAACCAATACTGACCCGCGCATCTCTGGCTTTGAGGCCACTTTTCTTCCAGCTG 1064
Db 2183 ACATGTATAACCAACTGAGCGCGGCATCTCCGGCTTTGAGGCCACTTTTCTTCCAGCTG 2124
Qy 1065 CCCAAGATGAGAGCTGTGGGGCTTTTGTGAGTGACACCCCAAGGGACATTTAGACGGCCC 1124
Db 2123 CTAAGATGAGAGCTGTGGGGCGCTTACGTAAAGCCCGAGGGGACATTCACACAGCCCC 2064
Qy 1125 TACTATCCAGGCCACTACCCGCCCAACATCAACTGCACATGGAATATCAAGGTGCCCAAC 1184
Db 2063 TACTATCCAGGCCACTACCCGCCCAACATTCGACTGTGACATGGAATATGAGGTGCCCAAC 2004
Qy 1185 AACCGAACTGAGGTGCGCTTCAAACTCTTCTATGTGTGACCCCAACGATACAGTGT 1244
Db 2003 AACCGAATGTGAGGTGAGCTTCAAACTCTTCTATGTGTGAGCGCGCTGCTGCTGCG 1944
Qy 1245 GGTCTCTGACCAAGGACTGTGGAGTCAACGGGGAGAGTACTGGGTGTGAGAGTTC 1304
Db 1943 GGCACCTTCCCGGAGGACTGTGGAGTCAATGGGGAGAAATACTGCGGAGAGAGTTC 1884
Qy 1305 CAGTTTGTGTGAGCAGCAACAGCAAGATACAGTCCACTTCCACTTCTGATCACTCG 1364
Db 1883 CAGTTTGTGTGAGCAGCAACAGCAACAGATCAAGTTCGCTTCCACTCAGATCACTCC 1824
Qy 1365 TACAGGACACCGGGTCTTGTAGTACTCTCTTACGACTTCCAGACGAGCCGTGCCCA 1424
Db 1823 TACAGGACACCGGGTCTTGTAGTACTCTCTTACGACTTCCAGTCCAGTCCCATGCCCC 1764
Qy 1425 GCGATGTTCATGTGCAAGACTGTGACAGTGCATCCGAAAGGAACTGCGCTGCGAGCGTGG 1484
Db 1763 GCGATGTTCATGTGCGGACCGGGCGGTGTATCCGGAAGGAGTGTGCGTGTGTGCGTGG 1704
Qy 1485 GCAGACTGCCCGGATATAGTGTGAGCGTGTACTGCGGATGCAATGCCACCCACAGTTC 1544
Db 1703 GCGGACTGACCGACACAGCAGATGAGTCAACTGCAAGTTGCGACGCGCGCCACAGTTC 1644
Qy 1545 ACGTGCAAAACAGTTCGCAAGCGCTCTTCTGGGTCTGTGACAGTGTCAAGACTGT 1604
Db 1643 ACGTGCAAAACAGTTCGCAAGCGCTCTTCTGGGTCTGTGCGACAGTGTGAACAGTGC 1584
Qy 1605 GGGGAGGAGTGCAGGAGGGCTGCGAGTGTCTGCTGGGAGTTTCAAGTGTTCAT 1664
Db 1583 GGGAGAACAGCAGCAGGAGGGGTGAGTGTGCGGCCCGAGACCTTTCAGGTGTTCAT 1524
Qy 1665 GGGAGGTGTCTCCCTCAGAGCCAGAGTGTATGGGAAGGACAACTGTGGAGATGGGTCT 1724

Db 1523 GGGAGTGGCTCTCGAAAGCCAGCAGTGAATGGGAAGACGACTGTGGGGAGCGGTCC 1464
Qy 1725 GAGAGGCTTCATGTGACAGCGTGAATGTCTCTTGGACCAATATATACCTACCGCTGC 1784
Db 1463 GAGAGGCTTCCTGCCCAAGGTGAACGTGTCTCACTTGTACCAACACACCTACCGCTGC 1404
Qy 1785 CAAAATGGCCCTCTGTCTGAGCAAGGGCAACCCCTGAGTGTGATGGGAAGAGAGAGCTGTAGC 1844
Db 1403 CTCAATGGGCTCTGTCTGAGCAAGGGCAACCCCTGAGTGTGACGGGAAGAGAGCTGTAGC 1344
Qy 1845 GATGGCTCCATGAGAAAACTGTGACTGTGGCTGGATCTCTTTTACCAACAGAGCTGCG 1904
Db 1343 GAGGGCTCAGATGAGAAAGACTGCACTGTGGCTGGCTGGCTCACTTACGAGACAGAGCTCGT 1284
Qy 1905 GTGGTTGGTGCGCAAGTACGAGAGGGAGTGGCCCTGGCAGGTGAGCTCCAGGCC 1964
Db 1283 GTTGTGGGGCAGGATCGGATGAGGGAGTGGCCCTGGCAGGTGAGCTTACGATGCT 1224
Qy 1965 CTGGGCCAGGCCACTTGTGTGGGGCTCGCTCATCTCTCTGACTGGCTGGTCTCTGCA 2024
Db 1223 CTGGGCCAGGCCACATCTCGGTGCTTCCCTCATCTCTCCAACTGGCTGGTCTCTGOC 1164
Qy 2025 GCTCATTTGCTTTAGGATGACAAAAATTTCAAGTACTCAGACTACAGATGTGAAGGCC 2084
Db 1163 GCACACTGCTACATCGATGACAGAGGATTCAGGTACTCAGACCCCAAGCTGGAAGGCC 1104
Qy 2085 TTCTGTGGTCTGTGGACCAAGCAAGCGAGTGGCTCTGGGGTGCAAGAGCTGAAGCTC 2144
Db 1103 TTCTGTGGTCTGACGACCAAGCGAGCGAGCGCCCTGGGGTGCAAGAGCGAGCTC 1044
Qy 2145 AAAGTATCATACCCACCTCTTCAATGATTTCACTTGGATGATGATGATGATGATGATG 2204
Db 1043 AAGCGCATCATCTCCCAACCCCTCTTCAATGATTTCACTTGGATGATGATGATGATG 984
Qy 2205 CTGAGCTGAGAGAGTGGGTGAGTACAGACCGTGTGGCGGCCCACTGCTGCTGCTGAT 2264
Db 983 CTGAGCTGAGAGAACCGGAGAGTACAGTCCATGTTGGCGGCCCACTGCTGCTGCGGAC 924
Qy 2265 GCTACCATGCTTCTTCCCTGTGGCAAGGCCATCTGGGTACAGGCTGGGGGCAACAAAA 2324
Db 923 GCCTCCCATGCTTCTTCCCTGGCGCAAGGCCATCTGGGTACAGGCTGGGGGCAACCCAG 864
Qy 2325 GAGGAGGTACCGAGCGCTGATCTCTGAGAGGTGAGATTCGTGTATCAACAGGAC 2384
Db 863 TATGAGGCGCATGCGCGCTGATCTCTCAAAAGGTGAGATTCGCGCTCATCAACAGGAC 804
Qy 2385 ACCTGTGAGGACCTCATGCGCGAGCAGATCACCCACGAATGATGTGTGGGTTTCTC 2444
Db 803 ACCTGCGAGAACCTTCTGCGCGAGCAGATCACGCCGCGCATGATGCGTGGGCTTCTC 744
Qy 2445 AGTGGGGTGTGACTCTCTGCAAGGTGACTTGTGTGGCCCTTGTCAAGCGGGAGAA 2504
Db 743 AGCGGCGCTGACTCTCTGCAAGGTGATTCGCGGGGACCCCTGTCCAGCGTGGAGCG 684
Qy 2505 GATGGGCGAATGTTCCAGGCTGTGTGTGAGCTGGGGTGAAGCTGCGCTCAGAGAAC 2564
Db 683 GATGGGCGAATCTTCCAGGCGGTGTGTGTGAGCTGGGGGAGACGCTGCGCTCAGAGAAC 624
Qy 2565 AAGCAGGCGGTGTACAAAGCTCCCTGTAGTTCGGGACTGGATCAAGAGACACTGGG 2624
Db 623 AAGCAGGCGGTGTACAAAGCTCCCTGTGTTCGGGACTGGATCAAGAGAACACTGGG 564
Qy 2625 GTATAGCAGCATGAGCAGACGCGACCAACCAACCCACAGGGATCCCGACATGCACA 2684
Db 563 GTATAGGCGCGCGG- -GCCACCCAAATGTGTACACCTGCGGGGCCACCCATGCTCCACC 507
Qy 2685 CCTGATACAGGAGAGAACACTGACGACATTTATGCTGTGGCTTCCCGCCCGCCCAACACA 2744
Db 506 CCAGTGTGCAG- CCTGAGGCTGGAGACTGGACCGCTGACTGACACGAGCGCC- CCAGA 449
Qy 2745 ACCCAGACTGTGAACCTGATCCTTAGGACTCAGAGT 2780

448 ACATACACTGTGAACCTCAATCTCCAGGGCTCCAAAT 413
RESULT 13
AX207897
LOCUS 3147 bp DNA linear PAT 31-AUG-2001
DEFINITION Sequence 1 from Patent WO0157194.
ACCESSION AX207897
VERSION AX207897.1 GI:15422495
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS Madison,B.L., Ong,B.O. and Yeh,J.C.
TITLE Nucleic acid molecules encoding transmembrane serine proteases, the
JOURNAL encoded proteins and methods based thereon
CORVAS PATENT: WO 0157194-A 1 09-AUG-2001;
INTERNATIONAL, INC. (US)
FEATURES
Location/Qualifiers
1..3147
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/notes="Nucleotide encoding MTSP1"
23..->2589
/notes="unnamed protein product"
/codon_start=1
/protein_id="CAC60377.1"
/db_xref="GI:15422496"
/translation="MGSDRRKGGGPKDFGAGLKYNSRHEKVNGLSEGVFLPVNNV
KVEKHGPRVWVLAALVLLGLLGLVHLYQDRVVRQKVFNGVMRIINENFV
DAYENSTFVSLASKVDALLKLYSGVFLGPHYKHSATVAFSGSVIAYVWSFVS
IQHIVLEABERVMABERVMLPRARSLAKSVVTSVAFPTDSTKVQRTQDNGSGFL
HARGVLEPTTGPEDSPYBAHARCOWALRGDADSVLSITRSPDLASDERGSDIV
TYNTLSFMEPHALVOLCGTTPSYNLTTHSSONVLLITLIIINTERHGFERTFQL
PMSCGGRLKRAQGTFFSPYYPHNPINDCTWNIENVNPNHVKVSFFLYLLEPGV
PRGCTPKDYVEINGEKYCGERSQFVVTSNSKITVRFHSDQSYDTGTFLAELYSYS
DPCPGQCTRTGCRKELRCWADCTDHSDELNCSDCAGHOFCKNFKCPKLVVVT
DSVNDGDNDSGCGSPACFCNSGKCLSKSQCCNGKDDCGDSDEASCPRNVVT
CTKHYRCLNGCLSKGNPECDGKEDCSGDEKDCDGLRSTROARVVVGTADDEG
EMQVSVHALGQGHSHIGASLISPNWLAAHCYIDDRFSDYDQFTQWTAFLGLHQS
APQVQVBRRLKRIISHPFNDFTDYDIALLELEKPAEYSVMVPICLPDASHVFP
AKGAIWVTQWGTQGTGALILQKGEIRVINTCENLLPQOITPRMVCVGFSLSGV
DSCQSDSGGLSLSSVEADGRIFQAGVGVSWGDGCAQRNKPQVYTRLPFRDWIKENTGV"
BASE COUNT 654 a 952 c 958 g 583 t
ORIGIN
Query Match 60.6%; Score 1883.2; DB 6; Length 3147;
Best Local Similarity 81.2%; Pred. No. 0;
Matches 2223; Conservative 0; Mismatches 508; Indels 5; Gaps 3;
Qy 45 GATCGACGCGCAAAACCATGGGTAGCAATCGGGCGCGCAAGCGCGAGGGGCTCTCAG 104
Db 5 GAGCGGCTCTGGGGTACCATGGGGAGCGATCGGGCCCGCAAGGGCGAGGGGCGCGAAG 64
Qy 105 GACTTCGGCGCGGAGACTCAAGTACAACTCCCGGCTAGAGAACATGAATGGCTTTGAGGAG 164
Db 65 GACTTCGGCGCGGAGACTCAAGTACAACTCCCGGCGACGAGAAAGTGAATGGCTTGGAGGA 124
Qy 165 GGTGTGGAGTTCCTGCTGCGAACCAATGCCAGAAATGGAGAGCGAGGCCCGCCAGCGC 224
Db 125 GCGCTGGAGTTCCTGCGCAGTCAACACAGTCAAGAGGTGAGAAAGCATGGCCCGGCGCGC 184
Qy 225 TGGGT 284
Db 185 TGGGT 244
Qy 285 CTGGTGTGGCAGTTCATTATCGGAATGTGGGGGTTCAAAAAGCTCTTCAATGGCCACTG 344
Db 245 CTGGTGTGGCAGTTCATTATCGGAATGTGGGGGAGTGGGTGTCAGAGAGTCTTCAATGGCTACATG 304

QY 345 AGGATCAAAATGAGATCTTTCTGATGCGTATGAGAACTCCACCTCCACAGAGTTTATC 404
DB |||||
QY 305 AGGATCAAAATGAGATCTTTCTGATGCGTATGAGAACTCCACCTCCACAGAGTTTATC 364
DB |||||
QY 405 AGCTGGCCAGCAGGTGAGAGGCGCTGAGCTGCTGATGAGTGAATGAGTCCCTGCTG 464
DB |||||
QY 365 AGCTGGCCAGCAGGTGAGAGGCGCTGAGCTGCTGATGAGTGAATGAGTCCCTGCTG 424
DB |||||
QY 465 GGTCCCTACCAAGAGTGGCTGTAATGCTGCTTCACTGAGGGGAGTGTATGCTCCCTAC 524
DB |||||
QY 425 GGCCTCTACCAAGAGTGGCTGCTGAGCGGCTTACAGGAGGAGCGTATGCTCCCTAC 484
DB |||||
QY 525 TACTGGTCAGAGTTCAAGATCCCTCCCAACCTGCGAGAGAGTTGATCGCGCCATGGCT 584
DB |||||
QY 485 TACTGGTCAGAGTTCAAGATCCCTCCCAACCTGCTGAGGAGGAGCGGCTCATGGCC 544
DB |||||
QY 585 GTGGAGCGAGTTGTAACATTTGCCACCCCGAGCAGCGGCACCTGAAATCTTGTGCTAACA 644
DB |||||
QY 545 GAGGAGCGAGTGTATGCTGCTCCCGCGGGCGGCTCCCTGAAATCTTGTGCTCACC 604
DB |||||
QY 645 TCTGTGGTGGCTTCCCAATTTGACCCCAAGATGCTGACAGAGACTCAGACAACAGCTGC 704
DB |||||
QY 605 TCAGTGGTGGCTTTCCTCCCAAGACTCCAAACAGTACAGAGAGCCAGACAACAGCTGC 664
DB |||||
QY 705 AGTTTGGCCCTGATGCTCCATGCTGAGAGTGTGATGCTGCTGCTGATGAGTGGAGTGCAC 764
DB |||||
QY 665 AGCTTTGGCTGACAGCGCCCGGGTGTGAGCTGTATGCTGCTTCAACACGCCCGCTTCCCT 724
DB |||||
QY 765 AACAGTCCCTACCGGCGCATGCGCGCTGCGAGTGGTCTGCGGGGAGCGCGACTCT 824
DB |||||
QY 725 GACAGCCCTTACCCCGCTATGCTCCCGCTGCGCAGTGGGCGCTTGGGGGGAGCGCGACTCA 784
DB |||||
QY 825 GTGCTGAGCCTCACCTTCCGAGCTTTGATGCTGCTGCTGCTGATGAGTGGAGTGCAC 884
DB |||||
QY 785 GTGCTGAGCCTCACCTTCCGAGCTTTGACCTTGGCTCTGCGACAGCGCGGAGCGAC 844
DB |||||
QY 885 CTGGTCACCGTGTATGATGAGCTGAGCCCGCATGGAACCCCGCTGCTGGTGGGCTGTGT 944
DB |||||
QY 845 CTGGTGAAGCTGTAAACACCTTGAAGCCCTGAGAGCCCGCTGAGAGCCCGCTGGTGGTGT 904
DB |||||
QY 945 GGCACCTTCTACCTCTTACAACTGACTTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1004
DB |||||
QY 905 GGCACCTTACCTCTCTTACAACTGACTTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 964
DB |||||
QY 1005 ACCTGTATTAACCAATPACTGACCGGCGCATCTGCTGCTTTGAGGCGACCTTTCTTCAGCTG 1064
DB |||||
QY 965 ACCTGTATTAACCAATPACTGACCGGCGCATCTGCTGCTTTGAGGCGACCTTTCTTCAGCTG 1024
DB |||||
QY 1065 CCCAAGATGAGAGCTGTGGCGGCTTTTGTAGTGACACCCAGGAGCAATTTAGAGCCCC 1124
DB |||||
QY 1025 CCTAGGATGAGAGCTGTGGAGCGGCTTACGTAAAGCCCGAGGAGCAATTAACAGAGCCCC 1084
DB |||||
QY 1125 TACTATCCAGGCGCACTACCCGCGCAACATCAACTGACATGGAATATCAAGGTGCCAAC 1184
DB |||||
QY 1085 TACTATCCAGGCGCACTACCCGCGCAACATCAACTGACATGGAATATCAAGGTGCCAAC 1144
DB |||||
QY 1185 AACCGGAACGTGAAGGTGCTTCAAACTCTTCTATCTGTGGAGCCCAAGCTACCAAGT 1244
DB |||||
QY 1145 AACAGCATGTGAAGTGAAGTCTCAAACTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1204
DB |||||
QY 1245 GGCTCTTGCAACAGGACTATGTGAGATCAACCGGGGAGAGTACTGCGGTGAGAGTCC 1304
DB |||||
QY 1205 GGCACCTCCCGCAAGGACTACGTGGAGATCAATGGGGAGAAATATCTGCGAGAGAGTCC 1264
DB |||||
QY 1305 CAGTTTGTGGTGAAGCAGCAACAGCAGCAAGATTACAGTCCACTTCTTCTTCTTCTTCTTCT 1364
DB |||||
QY 1265 CAGTTCGTGTCACAGCAACAGCAACAGATCAAGTTCGTTCCATCTAGATCAGTCC 1324
DB |||||
QY 1365 TACACGGAACCGGGTCTTCTAGCTGAGTACCTCTCTTACGACTTCCAAAGACCCCGTCCCA 1424
DB |||||
QY 1325 TACACGGAACCGGGTCTTCTAGCTGAATACCTCTCTTCTTCTTCTTCTTCTTCTTCTTCT 1384
DB |||||
QY 1425 GGGATGTTCAATGTGCAAGACTGGAGCGGTGCATCCGAAAGGAACCTGCGCTGCGAGCGGCTGG 1484
DB |||||

DB |||||
QY 1385 GGCAGTTTCAGTCCGCAAGGGCGGTATCCGAGAGGAGCTCGCTGTGATGGCTGG 1444
DB |||||
QY 1485 GCAGACTGCCGAGTATGATGATGAGGCTTATCTGCGCATGATGATGCAACCCACAGTTTC 1544
DB |||||
QY 1445 GCCAGTGCACCGACCAAGCGATGAGCTCAACTGCACTTGGCAGCGCGGCGCACAGTTTC 1504
DB |||||
QY 1545 ACGTGCAAAAACCAAGTTCTTCAAGCCCTCTTCTGGGTCTGTGACAGTGTCAACAGCTGT 1604
DB |||||
QY 1505 ACGTGCAAAAACCAAGTTCTTCAAGCCCTCTTCTGGGTCTGTGCAAGTGTGACAGCTGT 1564
DB |||||
QY 1605 GGGAGCGAAGTGAAGAGGAGGCTGCAAGCTGTCTTCTGGGAGTTTCAAGTGTTCCTCAAT 1664
DB |||||
QY 1565 GGGAGCAACAGCAGCAGCAGGAGGCTGCAAGTGTCTGCGGCCAGACCTTCAGGTGTTCCTCAAT 1624
DB |||||
QY 1665 GGGAGAGTGTCTCCCTCAGAGCAGAGTGAATGGAAAGGACAACTGTGGAGATGGGTCT 1724
DB |||||
QY 1625 GGGAGAGTGTCTCTCAGAAAGCCAGCAGTGAATGGAAAGGAGCAGCTGTGGGAGCGGCTCC 1684
DB |||||
QY 1725 GACGAGGCTTCATGTGACAGCGTGAATGTCTCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1784
DB |||||
QY 1685 GACGAGGCTCTCTGCCCCAAGTGAACGTCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1744
DB |||||
QY 1785 CAAATGAGCCTCTCTTCTGAGCAAGGCGCAACCTGAGTGTGATGGGAGAGCAGGACTGTAGC 1844
DB |||||
QY 1745 CTCAATGGGCTCTCTTCTGAGCAAGGCGCAACCTGAGTGTGACGGAAGGAGGACTGTAGC 1804
DB |||||
QY 1845 GATGGCTCCGATGAGAAACCTGTGACGTGGGCTGCGATCTCTTCAAAACAGGCTGCG 1904
DB |||||
QY 1805 GACGGCTCAGATGAGAAAGGACTGCGACTGTGGGCTGGGGTCAATTCACGAGACAGGCTCGT 1864
DB |||||
QY 1905 GTGGTGTGTCACGAATGCGAGCGAGCGAGTGGCCCTGCGAGGTGAGCTCCACAGCC 1964
DB |||||
QY 1865 GTTGTGGGGGCAAGATGCGGATGAGGGCGAGTGGCCCTGCGAGGTAAAGCTGCATGCT 1924
DB |||||
QY 1965 CTGGCCAGAGGCGCACTTGTGGGGCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2024
DB |||||
QY 1925 CTGGCCAGAGGCGCACTTGTGGGTGCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1984
DB |||||
QY 2025 GCTCATTTGCTTTCAGGATGACAAATTTCAAGTACTCAGACTACAGATGTGACAGGCG 2084
DB |||||
QY 1985 GCACACTGCTACATCGATGACAGAGGATTCAGGTACTCAGACCCCGCAGCTGAGAGCGCC 2044
DB |||||
QY 2085 TTCTGTGCTCTGTGGGCGCAGAGCAAGCGAGTGCCTCTGGGGTGCAGAGCTCAAGCTC 2144
DB |||||
QY 2045 TTCTGTGGCTTGCAGACCGAGCCAGCGCGCTTGGGGTGCAGAGCGAGGCTC 2104
DB |||||
QY 2145 AAACGTATCATCACCCCTTCTTCAATGATTTTCACTTTCGATATGACATCGCCTTG 2204
DB |||||
QY 2105 AAGCGCATCATCTCCACCCCTTCTTCAATGATTTCACTTTCGATATGACATCGCGCTG 2164
DB |||||
QY 2205 CTGGAGCTGGAGAGTGGTGGAGTACAGACCGCTGCTGCGCCCTCTCTCTCTCTCTCTCTCT 2264
DB |||||
QY 2165 CTGGAGCTGGAGAGAACCGCAGAGTACAGCTTCAATGGTGGCGGCTCTCTCTCTCTCTCT 2224
DB |||||
QY 2265 GCTACCCATGTCTTCTCTCTGTCGCAAGGCGCATCTGGGTCAAGGCTGGGGGACACAAAA 2324
DB |||||
QY 2225 GCCTCCCATGTCTTCTCTCTGCGCGCAAGGCGCATCTGGGTCAAGGCTGGGGGACACCCAG 2284
DB |||||
QY 2325 GAGGAGGTACCGAGAGCGCTGATCTCTGAGAGAGGTGAGATCCCTGTCTCATCAACAGAC 2384
DB |||||
QY 2285 TATGAGAGGCACTGGCGGCTGATCTCTGCAAAAGGAGTGGAGTCCCGCTCATCAACAGAC 2344
DB |||||
QY 2385 ACCTGTGAGAGACTCATGCGCAGCAGATCACCCCGCAGATGATGTGTGGGTCTCTCTCT 2444
DB |||||
QY 2345 ACCTGTGAGAGAACTTCTTCTGCGCAGCAGATCACCGCCGCGATGATGTGTGGGTCTCTCT 2404
DB |||||
QY 2445 AGTGGGGGTGGAGCTCTCTGCGAGGAGTCACTCTGGTGGCGCCCTTGTCAAGCGCGAGAAA 2504
DB |||||
QY 2405 AGCGCGCGGTGGAGTCTCTGCGAGGAGTTCGCGGGGAGCCCTGTCCAGCGGTGAGGCG 2464
DB |||||
QY 2505 GATGGGCGAATGTTTCCAGGCTGGTGGTGGAGCTGGGGTGAAGGCTGCGCTCAGAGAAC 2564
DB |||||

Db	2465	GATGGCGGAGTCTTCCAGGCGGTGTGGTGGAGACGGCTGCGCTCAGAGGAAC	2522
QY	2565	AAGCCAGGCGGTGTACACAAGGCTCCCTGTAGTTTCGGGACTTGGATCAAGAGCACACTGGG	2624
Db	2525	AAGCCAGGCGGTGTACACAAGGCTCCCTCTGTTTCGGGACTTGGATCAAGAGACACTGGG	2584
QY	2625	GTATAGCAGCATGGACNAGCAGCGGACCAACAACCCACAGGAGTCCCGACATGCACA	2684
Db	2585	GTATAGGGGCGGG--GCCACCCAAATGTGTACCTGCGGGGCCACCATGTCACC	2641
QY	2685	CCTGGATACAGGAGAGGAACACTGACACATTTATGCTGTGGCTCTCCCCCCCAACACA	2744
Db	2642	CCAGTGTGCAG-CCTCAGGCTGGAGACTGGACCGCTGACTGCACCGGCCC-CCAGA	2699
QY	2745	ACCAGACTGTGAAGTGCATCTTAGGACTCAGAGT	2780
Db	2700	ACATACACTGTGAAGTCAATCTCCAGGGCTCCAAAT	2735
RESULT 14			
LOCUS	AX207945	3147 bp	DNA linear PAT 31-AUG-2001
DEFINITION	Sequence 49 from Patent WO0157194.		
ACCESSION	AX207945		
VERSION	AX207945.1	GI:15422543	
KEYWORDS			
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
REFERENCE	1		
AUTHORS	Madison,E.L., Ong,E.O. and Yeh,J.C.		
TITLE	Nucleic acid molecules encoding transmembrane serine proteases, the encoded proteins and methods based thereon		
JOURNAL	Patent: WO 0157194-A 49 09-AUG-2001;		
	CORVAS INTERNATIONAL, INC. (US)		
FEATURES	Location/Qualifiers		
source	1..3147		
	/organism="Homo sapiens"		
	/mol_type="genomic DNA"		
	/db_xref="taxon:9606"		
CDS	<1865..2590		
	/note="protease domain of MTSPl"		
	/codon_start=1		
	/protein_id="CAC60383.1"		
	/db_xref="GI:15422544"		
	/translation="VVGGTDDADGEWPVWSLHAGGHICGSLISPNWLYSAACY		
	IDRGPRYSDPTQWTFALGHDSQSAFGVQRRLKRIISHFFPNDFDFTDYDIALLE		
	LEKPYISYRWRIELCPDASHVPAGKALIVTWGHTQYGGTGALLIQKEIRIVNOT		
	TCNPLPQITPTPRMVCVGLSGVSDSCQDGGPLSVSEADGRIFQAGVYVSWDGGCAQ		
	RNKPQYVTRPLPLFRDWIKENTGV"		
BASE COUNT	654 a 952 c 958 g 583 t		
ORIGIN			
Query Match	60.6%;	Score 1883.2;	DB 6; Length 3147;
Best Local Similarity	81.2%;	Pred. No. 0;	
Matches 2223;	Conservative	0; Mismatches 508;	Indels 5; Gaps 3;
QY	45	GATCGGACGCCAAAACCATGGGTAGCAATCCGGCCGACCGGCGGAGGGGCTCTCAG	104
Db	5	GAGCGGCTCGGGGTACCATGGGANGCATCGGGCCCGCAAGGCGGAGGGGCCCGAAG	64
QY	105	GACTTCGGCGCGGACTCAAGTACAACTCCCGCTAGAGAACATGAATGCTTTGAGGAG	164
Db	65	GACTTCGGCGCGGACTCAAGTACAACTCCCGCACAGAAAGTGAATGCTTGGAGAA	124
QY	165	GGTGTGAGTTCTGCTGCGAAACAATGCCAAGAAAGTGGAGAAGCGAGGCCCAAGCGC	224
Db	125	GGCGTGAGTTCTCGCCAGTCAACAAGGTGAAAAGCATGATGCCCGGGGCGC	184
QY	225	TGGTGTGCTGGTGGCAGTCTGTTTTCAGCTTCTTCTCTCCCTCATGCGTGGCTTG	284
Db	185	TGGTGTGCTGGCAGCGCTGATCGGCTCTCTTGTCTTGTCTTGTGGGATCGGCTTC	244

1365 TACAGGACACCGGGTCTTAGCTGAGTACCTCTCTACGACTCCAAACGACCGCTGCCA 1424
1375 TACAGGACACCGGGTCTTAGCTGAGTACCTCTCTACGACTCCAAACGACCGCTGCCA 1384
1425 GGGATGTTATGTGCAAGATGGAACGGTGCATCCGAAAGGAACCTGCGTGCAGCGCTGG 1484
1385 GGGCAGTTACGTCCGACACCGGGCGGTGTATCCGGAAGGAGCTGCGCTGTGATGGCTGG 1444
1485 GCAGACTGCCCGGATATAGTGTAGTGTAGTGTAGTGTAGTGTAGTGTAGTGTAGTGTAGT 1544
1445 GCCGACTGACCGACACGAGCGATGAGTCAACTGAGTGTGCGAGCGCGGACCAAGTTC 1504
1545 ACCTGCAAAAACGAGTGTCTGAAGCCCTCTCTGAGTGTGAGTGTGAGTGTGAGTGTGAGT 1604
1505 ACCTGCAAAAACGAGTGTCTGAAGCCCTCTCTGAGTGTGAGTGTGAGTGTGAGTGTGAGT 1564
1605 GGGGACGGAAGTGTGAGGAGGAGGCTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGT 1664
1565 GGGGACGGAAGTGTGAGGAGGAGGCTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGT 1624
1665 GGGGAGTGTCTCTCTGAGGAGGAGGAGGCTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGT 1724
1625 GGGGAGTGTCTCTCTGAGGAGGAGGAGGCTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGT 1684
1725 GAGGAGGCTTCATGTGACAGCGTGAATGTCTCTCTGAGGAGGAGGAGGCTGAGTGTGAGT 1784
1685 GAGGAGGCTTCATGTGACAGCGTGAATGTCTCTCTGAGGAGGAGGAGGCTGAGTGTGAGT 1744
1785 CAATGAGGCTCTCTGAGGAGGAGGAGGCTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGT 1844
1745 CTGAGGAGGCTCTCTGAGGAGGAGGAGGCTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGT 1804
1845 GATGGCTCCGATGAGGAGGAGGAGGCTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGT 1904
1805 GAGGAGGCTTCATGTGACAGCGTGAATGTCTCTCTGAGGAGGAGGAGGCTGAGTGTGAGT 1864
1905 GTGGTGTGAGGAGGAGGAGGAGGCTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGT 1964
1865 GTGGTGTGAGGAGGAGGAGGAGGCTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGT 1924
1965 CTGGGAGGAGGAGGAGGAGGAGGCTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGT 2024
1925 CTGGGAGGAGGAGGAGGAGGAGGCTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGT 1984
2025 GCTCATGTTTTCAGGATGACAAAATTTTCAAGTACTCAGACTACAGTGTGAGGAGGCTG 2084
1985 GCACACTGCTACATCGATGACAGAGGATTCAGTACTCAGACCCCAAGAGTGTGAGGAGG 2044
2085 TTCTGGGTCTGTGAGGAGGAGGAGGAGGCTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGT 2144
2045 TTCTGGGTCTGTGAGGAGGAGGAGGAGGCTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGT 2104
2145 AAAGGTATCATCACCCCTTCTCTCAATGATTTCACTTCACTTCACTTCACTTCACTTCACT 2204
2105 AAGGATCATCTCCACCCCTTCTCTCAATGATTTCACTTCACTTCACTTCACTTCACTTCACT 2164
2205 CTGAGGCTGAGGAGGAGGAGGAGGAGGCTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGT 2264
2165 CTGAGGCTGAGGAGGAGGAGGAGGAGGCTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGT 2224
2265 GCTACCATGCTTCTCTCTGAGGAGGAGGAGGAGGCTGAGTGTGAGTGTGAGTGTGAGTGTGAGT 2324
2225 GCTTCCATGCTTCTCTCTGAGGAGGAGGAGGAGGCTGAGTGTGAGTGTGAGTGTGAGTGTGAGT 2284
2325 GAGGAGGATACCGGAGGAGGAGGAGGAGGAGGCTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGT 2384
2285 TATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2344
2385 ACCTGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2444
2345 ACCTGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2404
2445 AGTGGGGGTGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2504

Db 2405 AGCGCGCGGTGACTCTCGCAGGGTGTTCGCGGGACCCCTGTTCAGCGTGGAGCG 2464
Qy 2505 GATGGCGAATGTTCCAGGCTGTGTGTGAGCTGGGGTGAAGCTGCGCTCAGAGGAAC 2564
Db 2465 GATGGCGGATCTTCAGGCGGTGTGTGAGCTGGGGAGACGCGCTGCGCTCAGAGGAAC 2524
Qy 2565 AAGCCAGCGGTGTACACAGGCTCCCTGTAGTTCGGGACTGGATCAAGAGCACACTGGG 2624
Db 2525 AAGCCAGCGGTGTACACAGGCTCCCTGTTCGGGACTGGATCAAGAGGAACACTGGG 2584
Qy 2625 GTATAGCAGCATGACAGACGAGCGACCAACACACACAGGAGTCCCGACATGCACA 2684
Db 2585 GTATAGGCGCGGG---GCCACCAATGTGTACACTTCGGGGCCACCATCTCTCCACC 2641
Qy 2685 CTGTGATACAGGAGAGGAACTGACGACATTTATGTGTGTGGCTTCCCGCCCCCAACACA 2744
Db 2642 CCAATGTGTGACG---CCTGAGGCTGGAGCTGACCGCTGACTGCACGAGCGCCC-CCAGA 2699
Qy 2745 ACCCAGACTGTGAGCTGATCCTCTTAGGACTCAGAGT 2780
Db 2700 ACATACACTGTGAACCTCAATCTCCAGGGCTCCAAAT 2735

RESULT 15
LOCUS AF057145 3147 bp mRNA linear PRI 22-SEP-2000
DEFINITION Homo sapiens serine protease TADG15 mRNA, complete cds.
ACCESSION AF057145
VERSION AF057145.1 GI:10257389
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 3147)
AUTHORS Tanimoto,H., Underwood,L.J., Wang,Y., Shigemasa,K., Parmley,T.H.
and O'Brien,T.J.
TITLE Direct Submission
JOURNAL Submitted (02-APR-1998) Biochemistry, University of Arkansas for
Medical Sciences, 4301 W. Markham St., Little Rock, AR 72205, USA
FEATURES
Location/Qualifiers
source 1..3147
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
23..2590
/notes="associated with human ovarian cancer"
/codon_start=1
/product="serine protease TADG15"
/protein_id="AAGI5395.1"
/db_xref="GI:10257390"
/translation="MGSDRARKGGGPKDFAGLKYNSRHEKYNGLKEGVEFLPVNV
KRYKEGPRGWRVLAALVGLLVLGLGFLVWLHQIQRDVRVQKRVNGYMRITNENPV
IPDHLVEAEERVNAERVMVLPKRLSLKSVFVTVAFPTDSKTQRTQDNCSFGL
HAGVLMPTFTTGPDPSPYPAHARQWALRGDASVLSLTPRSFSLASCDGSDGL
TVNTLSMPHEPHALVOLCGTYPSPYNTLTHSSQNVLLITLTNTRRHHPGFATFPL
PRMSSCGRLRKAQGFNSFYPGHYPPNIDCTNIEVENNQHVKSFFFLLEPGV
PAGCPDVEINGEKRSQFVVISNSKLTVPFSDQSTDTGFLAEYLSDS
DCPGQGTCTGTGIRKELCDGADCTDHSDELNCSDAGHQTCKNFKELPFWC
DSYNDGSDGDEQCPAQTPFCNSGKCLSKSQCNQKDDCGDSDGDEASCPKVVYT
CTKHYRLCGLSKNPECKDEKDCGLRSFTRQARVVGVDADBG
EWPQVSLHAIQGHICGASLISPMVLVSAACHVIDDRGPRYSDPTQWTAFLGHQDS
QRAPGVQERRLKIIISHPPFNDTFDYDIALLELEKPAEYSWVRPILCLPDSHPV
QKSAIWTGHTQYGTGALILQKREIRVINTTCENLLPQITPMCMCVGLSGSV
DSQSGGGLPSSVADGRIFQGVVSWGDCQARNKPGVYTRLPFRDIKENTGY"

BASE COUNT 654 a 952 c 958 g 583 t
ORIGIN

Query Match 60.6%; Score 1883.2; DB 9; Length 3147;
Best Local Similarity 81.2%; Pred. No. 0;
Matches 2223; Conservative 0; Mismatches 508; Indels 5; Gaps 3;

[illegible]

Search completed: February 21, 2004, 05:28:26
Job time : 11176 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 21, 2004, 02:03:32 ; Search time 795 Seconds
(without alignments)
10546.484 Million cell updates/sec

Title: US-09-900-751-1

Perfect score: 3106

Sequence: 1 catggtagacggctgccgg.....ttaaaaaaaaaaaaaaaaaa 3106

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 252756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq 19Jun03.*
1: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.*
2: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.*
3: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.*
4: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.*
5: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.*
6: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.*
7: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.*
8: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.*
9: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.*
10: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.*
11: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.*
12: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.*
13: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.*
14: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.*
15: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.*
16: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.*
17: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.*
18: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.*
19: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.*
20: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.*
21: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.*
22: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.*
23: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.*
24: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.*
25: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3106	100.0	3106	AAD37039	Epithin gene. Uni
2	1889.8	60.8	3149	AAA88493	Human matrixptase c
3	1884.8	60.7	3142	AAF28099	Human membrane-tyr
4	1883.2	60.6	3147	AAH87815	Tumour antigen der
5	1883.2	60.6	3147	AAD13113	Human membrane-tyr
6	1883.2	60.6	3147	AAD13155	Human membrane-tyr
7	1883.2	60.6	3147	AAH23601	Human TADG-15 codi
8	1883.2	60.6	3147	AAH23609	Human TADG-15 anti

9	1883.2	60.6	3147	24	AAI53444	Type II transmembr
10	1883.2	60.6	3147	24	AAI53445	Type II transmembr
11	1883.2	60.6	3147	25	ABZ58500	Transmembrane seri
12	1883.2	60.6	3147	25	ABZ58501	Transmembrane seri
13	1883.2	60.6	3147	25	ABZ22450	Human membrane-tyr
14	1883.2	60.6	3147	25	ABZ22451	Human MTSPI protea
15	1883.2	60.6	3147	25	AAD47180	Human membrane-tyr
16	1883.2	60.6	3147	25	AAD47181	Human membrane-tyr
17	1883.2	60.6	3147	25	AAD47225	Human membrane-tyr
18	1883.2	60.6	3152	22	AAH57431	Human intestine ce
19	1872.8	60.3	3159	21	AAH37657	Human peptidase, H
20	1793.8	57.8	3115	22	ABA08672	Human membrane-tyr
21	1793.8	57.8	3115	22	AAH9569	Human protein enco
22	1762.2	56.7	2955	21	AAA88492	Human matrixptase (
23	1559.2	50.2	2756	21	AAC77957	Human cancer assoc
24	1532.4	49.3	3112	24	ABH76526	CDNA encoding huma
25	1530.8	49.3	2900	22	AAH23602	Human SNC-19 codin
26	1498.8	48.3	3413	23	AAH85629	Human cDNA differe
27	1498.8	48.3	3413	23	AAH85629	DNA encoding novel
28	701.2	22.6	1553	24	ABK30240	Human G-protein-co
29	556.4	17.9	726	24	AAD36928	Human matrixptase o
30	551.8	17.8	723	18	AAH79128	Human marine prote
31	379.4	12.2	2086	23	AAH85628	DNA encoding novel
32	367	11.8	1458	23	AAH85625	DNA encoding novel
33	306.8	9.9	434	24	ABK39442	DNA encoding lung
34	306.8	9.9	434	25	ACA11771	Human lung cancer
35	306.8	9.9	434	25	ACA02957	Lung cancer therap
36	302.6	9.7	429	24	ABL82332	Human ovarian canc
37	279.2	9.0	387	23	AAH85626	DNA encoding novel
38	265	8.5	591	23	AAH88827	CDNA #1503 encodin
39	246.8	7.9	393	25	ABH40789	Bovine EST associa
40	242.4	7.8	2801	24	ABQ61196	Human PRO618 encod
41	240.8	7.8	2672	22	AAD13117	Human membrane-tyr
42	240.8	7.8	3104	22	AAD13116	Human membrane-tyr
43	239.2	7.7	1327	20	AAZ34034	Human EST DNA35597
44	239.2	7.7	1327	21	AAH78495	Human EST DNA35597
45	239.2	7.7	1327	21	AAH58237	Human EST (express

ALIGNMENTS

RESULT 1
AAD37039
ID AAD37039 standard; DNA; 3106 BP.
AC AAD37039;
XX
XX
XX 21-AUG-2002 (first entry)
XX Epithin gene.
XX Transgenic; transgenic animal; pharmacological therapy; gene therapy;
XX phenotype modulation; genetic disease; epithin; gene; ds.
XX Unidentified.
XX
XX Key Location/Qualifiers
XX CDS 63...2630
FT FT /*tag= a
FT FT /product= "Epithin protein"
FT FT 2332...2465
FT FT /*tag= b
FT FT /note= "Sequence flanking Neo insert in
FT FT targeting construct"
FT FT 2466...2465
FT FT /*tag= c
FT FT /note= "Sequence deleted in targeting construct"
FT FT 2466...3033
FT FT /*tag= d
FT FT /note= "Sequence flanking Neo insert in
FT FT targeting construct"
XX

PN WO200203787-A2.
 XX 17-JAN-2002.
 XX 06-JUL-2001; 2001WO-US21427.
 XX 06-JUL-2000; 2000US-216109P.
 PR 06-JUL-2000; 2000US-216251P.
 PR 06-JUL-2000; 2000US-216258P.
 PR 06-JUL-2000; 2000US-216768P.
 PR 10-JUL-2000; 2000US-217449P.
 PR 10-JUL-2000; 2000US-217450P.
 PR 10-JUL-2000; 2000US-217660P.
 PR 27-JUL-2000; 2000US-221491P.
 PR 27-JUL-2000; 2000US-221669P.
 PR 07-AUG-2000; 2000US-221670P.
 PR 07-AUG-2000; 2000US-223170P.
 PR 07-AUG-2000; 2000US-223172P.
 PR 07-AUG-2000; 2000US-223460P.
 PR 26-OCT-2000; 2000US-244037P.
 PR 26-OCT-2000; 2000US-244111P.
 PR 26-JUN-2001; 2001US-301217P.
 XX (DELTA-) DELTAGEN INC.
 PA Allen KD, Leviten MW;
 XX P-PSDB; AAE23083.
 PI WPI; 2002-154853/20.
 DR Novel non-human transgenic animal, preferably transgenic mice
 DR comprising disruption in target gene, e.g., trypsin gene, useful for
 DR identifying an agent that modulates expression or function of target
 PT gene.
 PT
 XX
 PS Example 2; Fig 2A; 74pp; English.
 XX
 CC The present invention relates to non-human transgenic animals preferably
 CC transgenic mice comprising disruption in target gene such as trypsin
 CC gene. The invention also relates to compositions and methods relating
 CC to the characterisation of gene functions. The transgenic animals are
 CC useful for identifying an agent that modulates the expression or function
 CC of a target. They are useful for identifying an agent that modulates a
 CC phenotype associated with a disruption in trypsin genes or in trypsin
 CC clotting factor protease-like genes by administering an agent to the
 CC transgenic animal and determining whether the agent modulates the
 CC phenotype where the agent has effect on decreased body weight, decreased
 CC thymus weight, decreased thymus to body weight ratio, increased pre-pulse
 CC inhibition, significant decrease in their response latency to the hot
 CC plate test or a decreased response threshold to metrazol. Agents that
 CC modulate the expression, function or activity of the target gene are
 CC useful for treating a disorder associated with a mutation in trypsin
 CC gene or in trypsin clotting factor protease-like gene. The transgenic
 CC animals are useful for testing the efficacy of proposed genetic and
 CC pharmacological therapies for human genetic diseases. They are useful
 CC as models for diseases, disorders or conditions associated with
 CC phenotypes relating to a disruption in a target and to identify drugs,
 CC pharmaceuticals, therapies and interventions which may be effective in
 CC treating a disease or other phenotypic characteristics of the animal.
 CC The present sequence is epitope gene. This sequence is used in the
 CC exemplification of the invention.
 XX
 SQ Sequence 3106 BP; 692 A; 862 C; 897 G; 655 T; 0 other;
 Query Match 100.0%; Score 3106; DB 24; Length 3106;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 3106; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 1 CATGGTAGACGGCTGCCCGGAGGACCAACCGCTGTGAGACCGGGATCGGACCGCAAAA 60
 Db 1 CATGGTAGACGGCTGCCCGGAGGACCAACCGCTGTGAGACCGGGATCGGACCGCAAAA 60
 Qy 61 CCATGGGTAGCAATCGGGGCGCGAAGCGCGGGGGCTCTCAGGACTTCGGCGGGGAC 120

Db 61 CCATGGGTAGCAATCGGGGCGCGAAGCGCGGGGGCTCTCAGGACTTCGGCGGGGAC 120
 Qy 121 TCAAGTACAACTCCCGGCTAGAGAAATGAAATGGCTTTTGGAGGGGTGTGGAGTTCTCTGC 180
 Db 121 TCAAGTACAACTCCCGGCTAGAGAAATGAAATGGCTTTTGGAGGGGTGTGGAGTTCTCTGC 180
 Qy 181 CTGCGAAATCCCAAGAAAGTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 240
 Db 181 CTGCGAAATCCCAAGAAAGTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 240
 Qy 241 CAGTGTCTGTTCAGCTTCT 300
 Db 241 CAGTGTCTGTTCAGCTTCT 300
 Qy 301 ATTATCGGAATGTGGGGTTCAAAAAGTCTTCAATGGCCATCTGAGGATCACAAAATGAGA 360
 Db 301 ATTATCGGAATGTGGGGTTCAAAAAGTCTTCAATGGCCATCTGAGGATCACAAAATGAGA 360
 Qy 361 TCTTTCTGATGCGTATGAGAACTCCACTCCACAGAGTTTATCAGCTTGGCCAGCAGG 420
 Db 361 TCTTTCTGATGCGTATGAGAACTCCACTCCACAGAGTTTATCAGCTTGGCCAGCAGG 420
 Qy 421 TGAAGGAGCGCTGAAGCTGTGTAATGAAGTCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 480
 Db 421 TGAAGGAGCGCTGAAGCTGTGTAATGAAGTCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 480
 Qy 481 AGTCGGCTGTAACTGCCTTTCAGTGGGCGAGTGTCTATCGCTTACTACTCTGTCAGAGTTCA 540
 Db 481 AGTCGGCTGTAACTGCCTTTCAGTGGGCGAGTGTCTATCGCTTACTACTCTGTCAGAGTTCA 540
 Qy 541 GCATCCCCCACTCTGGGAGAGAGTTGATCGGCCATCTGCTGTGAGAGAGAGAGAGAGAGAG 600
 Db 541 GCATCCCCCACTCTGGGAGAGAGTTGATCGGCCATCTGCTGTGAGAGAGAGAGAGAGAGAG 600
 Qy 601 CATTCGCAACCGGAGCAGCGGCACTGAAATCTTCTGTGTCTTAACTCTGTGTGTGTGTGTGTGT 660
 Db 601 CATTCGCAACCGGAGCAGCGGCACTGAAATCTTCTGTGTGTGTGTGTGTGTGTGTGTGTGT 660
 Qy 661 CCATTGACCCCAAGATGTGTCAGAGACTCAGGACAAAGCTGACGTTTGGCTTCCATG 720
 Db 661 CCATTGACCCCAAGATGTGTCAGAGACTCAGGACAAAGCTGACGTTTGGCTTCCATG 720
 Qy 721 CCATGGTGCAGAGTGCACGCTTCACTACCCCTGGCTTCCCAACAGCTTCCCTACCCG 780
 Db 721 CCATGGTGCAGAGTGCACGCTTCACTACCCCTGGCTTCCCAACAGCTTCCCTACCCG 780
 Qy 781 CGCATGCCCGCTGCCAGTGGGTCTCTCGGGGGAGCGCCGACTCTGTGTGTGTGTGTGTGTGTGT 840
 Db 781 CGCATGCCCGCTGCCAGTGGGTCTCTCGGGGGAGCGCCGACTCTGTGTGTGTGTGTGTGTGTGT 840
 Qy 841 TCCGAAGCTTTGATGTGCTTCCCTGTGATGAGATGGGAGTGCAGTGTGTGTGTGTGTGTGTGT 900
 Db 841 TCCGAAGCTTTGATGTGCTTCCCTGTGATGAGATGGGAGTGCAGTGTGTGTGTGTGTGTGTGT 900
 Qy 901 ATAGCTGAGCCCCCAAG 960
 Db 901 ATAGCTGAGCCCCCAAG 960
 Qy 961 CCTACAACTGACTTTCT 1020
 Db 961 CCTACAACTGACTTTCT 1020
 Qy 1021 CTGACCGGGAGATCTCTGGCTTGGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1080
 Db 1021 CTGACCGGGAGATCTCTGGCTTGGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1080
 Qy 1081 GTGGCGGCTTTTGTAGTGACACCCAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1140
 Db 1081 GTGGCGGCTTTTGTAGTGACACCCAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1140
 Qy 1141 ACCCGGCCCAACATCAATGCAATGAAATATCAAGTGTGCGGAGAGAGAGAGAGAGAGAGAGAG 1200

Db 1141 ACCGCCCAACATCAACTGCACATGGAATATCAAGGTGCCCAACAACCGGAAACGTGAAGG 1200
Qy 1201 TGGCTTTCAAACCTCTTATCTGGTGACCCCAACGTACAGTGGGCTCTGTGACCAAGG 1260
Db 1201 TGGCTTTCAAACCTCTTATCTGGTGACCCCAACGTACAGTGGGCTCTGTGACCAAGG 1260
Qy 1261 ACTATGTGGAGATCAACGGGAGAAAGTACTGCGGTGAGAGTCCAGTTTGTGTGAGCA 1320
Db 1261 ACTATGTGGAGATCAACGGGAGAAAGTACTGCGGTGAGAGTCCAGTTTGTGTGAGCA 1320
Qy 1321 GCAACAGCAGCAAGATTACAGTCACTTCATCTGATCATCTGTACACGACACCGGGT 1380
Db 1321 GCAACAGCAGCAAGATTACAGTCACTTCATCTGATCATCTGTACACGACACCGGGT 1380
Qy 1381 TCCTAGCTGAGTACTCTCTTACGACTCCAAACGACCGGTGCCAGGAGATGTTATGTGCA 1440
Db 1381 TCCTAGCTGAGTACTCTCTTACGACTCCAAACGACCGGTGCCAGGAGATGTTATGTGCA 1440
Qy 1441 AGACTGACGGTGCATCCGAAAGAACTGGCGTCCGACGGCTGGGACAGTGCCTCGGATT 1500
Db 1441 AGACTGACGGTGCATCCGAAAGAACTGGCGTCCGACGGCTGGGACAGTGCCTCGGATT 1500
Qy 1501 ATAGTGATGAGCGTTACTGCGGATGCAATGCCACCCACCGATTACGTGCAAAACCACT 1560
Db 1501 ATAGTGATGAGCGTTACTGCGGATGCAATGCCACCCACCGATTACGTGCAAAACCACT 1560
Qy 1561 TCTGCAAGCCCTCTTCTGGGTCTGTGACAGTGTCAACGACTGTGGGACGGAAGTGAAG 1620
Db 1561 TCTGCAAGCCCTCTTCTGGGTCTGTGACAGTGTCAACGACTGTGGGACGGAAGTGAAG 1620
Qy 1621 AGAGGCTGAGCTGCTCTGCTGGAGTTCAAGTGTCCAAATGGGAAAGTGTCTCCCTC 1680
Db 1621 AGAGGCTGAGCTGCTCTGCTGGAGTTCAAGTGTCCAAATGGGAAAGTGTCTCCCTC 1680
Qy 1681 AGAGGCTGAGCTGCTCTGCTGGAGTTCAAGTGTCCAAATGGGAAAGTGTCTCCCTC 1740
Db 1681 AGAGGCTGAGCTGCTCTGCTGGAGTTCAAGTGTCCAAATGGGAAAGTGTCTCCCTC 1740
Qy 1741 ACAGCGTGAATGTCTCTTGTGACCAAAATATCTACCGCTGCCAAATGGCTCTGTC 1800
Db 1741 ACAGCGTGAATGTCTCTTGTGACCAAAATATCTACCGCTGCCAAATGGCTCTGTC 1800
Qy 1801 TGACCAAGGCAACCTGATGTGATGGGAAGACGACTGTAGCGATGGCTCCGATCAGA 1860
Db 1801 TGACCAAGGCAACCTGATGTGATGGGAAGACGACTGTAGCGATGGCTCCGATCAGA 1860
Qy 1861 AAAAAGTGTGAGTGGCTGGATCTCTTACCAACAGGCTCGGCTGGTGGGACGA 1920
Db 1861 AAAAAGTGTGAGTGGCTGGATCTCTTACCAACAGGCTCGGCTGGTGGGACGA 1920
Qy 1921 ATCGGACGAGGCGAGTGGCTCTGCGAGGTGAGCTCCAGCCCTGGGCGCAGGCGCACT 1980
Db 1921 ATCGGACGAGGCGAGTGGCTCTGCGAGGTGAGCTCCAGCCCTGGGCGCAGGCGCACT 1980
Qy 1981 TGTGTGGGCTCGCTCATCTCTGCTGAGTGGCTGGCTGTGAGCTCATCTGCTTCAGG 2040
Db 1981 TGTGTGGGCTCGCTCATCTCTGCTGAGTGGCTGGCTGTGAGCTCATCTGCTTCAGG 2040
Qy 2041 ATGACAAAATTTCAAGTACTCAGATACAGATGTGAGCGGCTCTGCTGCTGCTGG 2100
Db 2041 ATGACAAAATTTCAAGTACTCAGATACAGATGTGAGCGGCTCTGCTGCTGCTGG 2100
Qy 2101 ACCAGACGAGCGAGTGGCTCTGGGCTGAGGAGTGAAGCTCAAAACGATCATCAACC 2160
Db 2101 ACCAGACGAGCGAGTGGCTCTGGGCTGAGGAGTGAAGCTCAAAACGATCATCAACC 2160
Qy 2161 ACCCTTCTCAATGATTTACCTTGCATATGATCATCGCTTCTGCTGAGCTGAGAAAT 2220
Db 2161 ACCCTTCTCAATGATTTACCTTGCATATGATCATCGCTTCTGCTGAGCTGAGAAAT 2220
Qy 2221 CGGTGGAGTACAGACCGCTGCTGGCGCCCATCTGCTGCTGCTGCTGCTGCTGCTTCC 2280
Db 2221 CGGTGGAGTACAGACCGCTGCTGGCGCCCATCTGCTGCTGCTGCTGCTGCTGCTTCC 2280

Qy 2281 CTGCTGSCAAGGCCATCTGGGTCAAGGCTGGGGGCACACAAAAGAGGGAGGTACCCGAG 2340
Db 2281 CTGCTGSCAAGGCCATCTGGGTCAAGGCTGGGGGCACACAAAAGAGGGAGGTACCCGAG 2340
Qy 2341 CGCTGATCTCTGSCAAGGGGTGAGATCCGTGTCTATCAACACGAGACCACTGTGAGACCTCA 2400
Db 2341 CGCTGATCTCTGSCAAGGGGTGAGATCCGTGTCTATCAACACGAGACCACTGTGAGACCTCA 2400
Qy 2401 TGGCGACGAGATCAACCCACGAAATGATGTGTGTGGTTCCTCAGTGGGGGTGTGACT 2460
Db 2401 TGGCGACGAGATCAACCCACGAAATGATGTGTGTGGTTCCTCAGTGGGGGTGTGACT 2460
Qy 2461 CCTGCCAAGGCTGACTCTGTGGTGGCCCTTGTCAAGCGCGGAGAAAGATGGGCGAATGTTCC 2520
Db 2461 CCTGCCAAGGCTGACTCTGTGGTGGCCCTTGTCAAGCGCGGAGAAAGATGGGCGAATGTTCC 2520
Qy 2521 AGCTGTGTGTGTGAGTGGGTGCAAGGCTGCGCTCAGAGAGAAACAAGCCAGGCGTGTACA 2580
Db 2521 AGCTGTGTGTGTGAGTGGGTGCAAGGCTGCGCTCAGAGAGAAACAAGCCAGGCGTGTACA 2580
Qy 2581 CAAGGCTCCCTGTAGTTCGGGACTGGATCAAGAGACACACTGGGGTATAGCAGCATGAC 2640
Db 2581 CAAGGCTCCCTGTAGTTCGGGACTGGATCAAGAGACACACTGGGGTATAGCAGCATGAC 2640
Qy 2641 AGACAGCCGACCAACCAACACCCAGGATGCCCGACATGCACACCTGGGATACAGAGAG 2700
Db 2641 AGACAGCCGACCAACCAACACCCAGGATGCCCGACATGCACACCTGGGATACAGAGAG 2700
Qy 2701 GAACACTGACGACATTTATGCTGTGGCTCCCCCCCCCAACACCAACCACTGTGAACT 2760
Db 2701 GAACACTGACGACATTTATGCTGTGGCTCCCCCCCCCAACACCAACCACTGTGAACT 2760
Qy 2761 GCATCTTAGGACTCAGAGTTCCTCAAAAGTGGGACCCCTCAAGAGTTGGAGAGAGAAC 2820
Db 2761 GCATCTTAGGACTCAGAGTTCCTCAAAAGTGGGACCCCTCAAGAGTTGGAGAGAGAAC 2820
Qy 2821 TTGCGTGTAGCGGCCAGCTGGGGGCAAGGCTTGTATGGGAGCCCTTCCCTCTAGCC 2880
Db 2821 TTGCGTGTAGCGGCCAGCTGGGGGCAAGGCTTGTATGGGAGCCCTTCCCTCTAGCC 2880
Qy 2881 CTGAGCTGGGTGAAGATGATGTGTCGGGAGAGTGTCTTCCAACTGTCTATTGAGCTCCC 2940
Db 2881 CTGAGCTGGGTGAAGATGATGTGTCGGGAGAGTGTCTTCCAACTGTCTATTGAGCTCCC 2940
Qy 2941 GGGAGCCCTATGGGAGAGGGGCTCAGGCTCACTCTTTTCCAGGAGCGCCAGCCCTAGGA 3000
Db 2941 GGGAGCCCTATGGGAGAGGGGCTCAGGCTCACTCTTTTCCAGGAGCGCCAGCCCTAGGA 3000
Qy 3001 ACCCCAGAAAAGAGTGGTACTAAGGCTGAAATTTGTTTCTGTTGCGAGGGGTGGGTAT 3060
Db 3001 ACCCCAGAAAAGAGTGGTACTAAGGCTGAAATTTGTTTCTGTTGCGAGGGGTGGGTAT 3060
Qy 3061 TTGAGAGTAAACATTTTATTTCTTTTAAAAAATAAAAAAATAAAAAA 3106
Db 3061 TTGAGAGTAAACATTTTATTTCTTTTAAAAAATAAAAAAATAAAAAA 3106

RESULT 2

AAA88493

ID AAA88493 standard; cDNA; 3149 BP.

XX AAA88493;

XX AC

XX XX

DT 22-JAN-2001 (first entry)

XX Human matrixase cDNA.

DE

XX

XX

KW Matrixase; serine protease; human; Barrett's epithelium;

KW actinic keratosis; leukoplakia; Barrett's epithelium;

KW columnar metaplasia; ulcerative colitis; Bowenoid papulosis;

KW adenomatous colorectal polyp; Qyemat erythroplasia;

KW vulvar intraepithelial neoplasia; tumour; metastasis; therapy; ss.

XX	OS	Homo sapiens.		QY	230	GGTGTGCTGGCAGTGTCTTCTCAGTCTTCTCTCCCTCATGGCTGGCTTGTCTGT	289
XX	XX			DB	203	GGTGTGCGCAGCCGTGTGATCGGCTCTCTCTTGTCTTGTGGGATCGGCTTCTGT	262
XX	FT	Location/Qualifiers		QY	290	GTGGCATTCTCCATTTATCGAATGTGGGTTCAAAAAGTCTTCAATGGCCATCTGAGGAT	349
XX	FT	36..2603		DB	263	GTGGCATTTCAGTACCGGACGTGTCTCAGAAAGGTCTTCAATGGCTATACAGGAT	322
XX	PN	/*tag= c		QY	350	CACAAATGAGATCTTTCTGGATGCGTATGAGAACTCCACCTCCACAGATTTATCAGCT	409
XX	PD	WO200053232-A1.		DB	323	CACAAATGAGAAATTTTGTGGATGCGTACGAGAACTCCAACTCCATGAGTTTGTAAAGCT	382
XX	PP	14-SEP-2000.		QY	410	GGCCAGCCAGGTGAAGGAGGGCGCTCAAGCTGTGTACAAATGAAGTCCCTGTCTGGGTCC	469
XX	PR	10-MAR-2000; 2000WO-US06111.		DB	383	GGCCAGCAAGGTGAAGGAGCGCTGAAGCTGTGTACAGGGAGTCCCATTCCTGGGCC	442
XX	XX	12-MAR-1999; 99US-0124006.		QY	470	CTACCAAGAAAGTCTGGCTGTAACTGTCTTCACTGAGGGCAGTGTCTATCGCTTACTGT	529
XX	PA	(GEO) UNIV GEORGETOWN.		DB	443	CTACCAAGAAAGTCTGGCTGTGACGGCTTCAAGCGGCTTCAAGCGAGGCGAGCTCATCGCTACTGT	502
XX	PI	Dickson RB, Lin C, Johnson M, Wang S, Enyedy I;		QY	530	GTGAGAGTTTCAAGTATCCCGCACCTCGCAGAGAGGTTGATCGCGCATGGCTGTGGA	589
XX	DR	WPI; 2000-594268/56.		DB	503	GTGAGAGTTTCAAGTATCCCGCACCTCGCAGAGAGGTTGATCGCGCATGGCTGTGGA	562
XX	DR	F-PSDB; AAB19552.		QY	590	GGGAGTTTCAATTTGCCACCCCGGAGCAGCGGCACTGAAATCCTTCGTCTTAACTCTGT	649
XX	PT	Treating malignancies, premalignant and pathologic conditions in a		DB	563	GGGAGTTTCAATTTGCCACCCCGGAGCAGCGGCACTGAAATCCTTCGTCTTAACTCTGT	622
XX	PT	subject, comprises administering matriptase modulating agent		QY	650	GGTGGCTTCCCATTTGAGCCCGAGAAATCTGTCAGAGGACTCTAGGACAAACAGCTGAGTTT	709
XX	PS	Claim 9; Fig 15; 116pp; English.		DB	623	GGTGGCTTCCCATTTGAGCCCGAGAAATCTGTCAGAGGACTCTAGGACAAACAGCTGAGTTT	682
XX	CC	The present sequence is that of cDNA coding for the full-length		QY	710	TGCGCTGTATGCCATGTGTGAGCAGTGTGAGCTGATGCGCTTCCACCGCCCGCTTCCCTG	769
XX	CC	form (see AAB19521) of human matriptase, a trypsin-like protease.		DB	683	TGCGCTGTATGCCATGTGTGAGCAGTGTGAGCTGATGCGCTTCCACCGCCCGCTTCCCTG	742
XX	CC	CDNA encoding a truncated form of matriptase is given in AAB8492.		QY	770	TGCGCTGTATGCCATGTGTGAGCAGTGTGAGCTGATGCGCTTCCACCGCCCGCTTCCCTG	829
XX	CC	Either form can be incorporated into a vector and used in a claimed		DB	743	CCGCTTACCCCGCTCATGCCGCTGTCAGTGTGGCCCTGCGGGGAGCGCCGACTCAGTGT	802
XX	CC	method for making recombinant matriptase. The zymogen (inactive)		QY	830	GAGCTTCACTTCCGAAAGCTTGTATGTCTCTCTGTGATGAGCATGGCAGTGTCTGT	889
XX	CC	form of matriptase is a single-chain protein. The active 2-chain		DB	803	GAGCTTCACTTCCGAAAGCTTGTATGTCTCTCTGTGATGAGCATGGCAGTGTCTGT	862
XX	CC	form strongly interacts with fragments of a Kunitz-type serine		QY	890	CACCTGTATGATAGCTTGGAGCCCATGAGCCCATGAGCCCATGAGCCCATGAGCCCATG	949
XX	CC	protease inhibitor (hepatocyte growth factor activator inhibitor,		DB	863	GACGCTGTATGATAGCTTGGAGCCCATGAGCCCATGAGCCCATGAGCCCATGAGCCCATG	922
XX	CC	HAI-1) to form SDS-stable complexes. In breast cancer cells,		QY	950	CTTCTCAACCTTCAACAACTGACTTCTCTCTCTCCAGAAAGTCTTCTCTCTCTCTCTCT	1009
XX	CC	matriptase is present mainly as the uncomplexed form. Only the		DB	923	CTTCTCAACCTTCAACAACTGACTTCTCTCTCTCCAGAAAGTCTTCTCTCTCTCTCTCT	982
XX	CC	complexed matriptase in detected in human milk. The invention is		QY	1010	GATACCAATCTGACCGGCGACATCTGCTGCTTGTGAGGCGCTTCTCTCTCTCTCTCTCT	1069
XX	CC	directed to a method of detecting a malignancy or a pre-malignant		DB	983	GATAACCAACATGAGCGGCGGCACTCCCGGCTTGTAGGCGCACCTTCTCTCTCTCTCTCT	1042
XX	CC	lesion in breast or other tissue by detecting the presence of		QY	1070	GATGAGCAGCTGTGGCGGCTTTTTCAGTGACACCCCAAGGAGCATTTAGCAGCCCTCTCT	1129
XX	CC	single- or 2-chain forms of matriptase in the tissue. The		DB	1043	GATGAGCAGCTGTGGCGGCGCTTTCAGTGACACCCCAAGGAGCATTTAGCAGCCCTCTCT	1102
XX	CC	object is to inhibit tumour onset, tumour growth and metastasis.		QY	1130	TCCAGGCCACTACCGGCCCAACATCACTGACATGGAATATCAAGGTGCCCAACACCG	1189
XX	CC	Malignancies and pre-malignant conditions characterised by		DB	1103	CCCAGGCCACTACCGGCCCAACATCACTGACATGGAATATCAAGGTGCCCAACACCG	1162
XX	CC	expression of the zymogen or activated form of matriptase are		QY	1190	GAACTGTGAGGTGGCTTCAAACTCTTCTATCTGTGTGACCCCAACAGTACAGTGGGCTC	1249
XX	CC	treated by administering an inhibitor of matriptase, especially a		DB	1163	GATGAGGAGTGTGGCGGCTTCAAACTCTTCTATCTGTGTGACCCCAACAGTACAGTGGG	1222
XX	CC	Bowman-Birk inhibitor. The pre-malignant condition is atypical		QY	1250	CTGCACCAAGGACTATGTGGAGATCAACCGGAGAAATGACTGCGGTGAGAGGTCCAGTT	1309
XX	CC	ductal hyperplasia of the breast, actinic keratosis, leukoplakia,		DB	1223	CTGCCCCAAGGACTATGTGGAGATCAATGGGAGAAATGACTGCGGAGAGAGGTCCAGTT	1282
XX	CC	Barrett's epithelium of the esophagus, ulcerative colitis,		QY	1310	TGTGTGAGCAGCAACAGCAGCAAGATTTACAGTCCACTTTCATCTCTGATCACTCGTAC	1369
XX	CC	adenomatous colorectal polyps, erythroplasia of the Queyrat,					
XX	CC	Bowen's disease, Bowenoid papulosis, vulvar intraepithelial					
XX	CC	neoplasia or dysplastic changes to the cervix. The invention also					
XX	CC	provides methods for in vivo or in vitro diagnosis of malignancy					
XX	CC	or pre-malignant lesion, and methods of identifying matriptase					
XX	CC	modulators, including activators and inhibitors.					
XX	XX						
XX	XX	Sequence 3149 BP; 647 A; 959 C; 962 G; 581 T; 0 other;					
XX	XX	Query Match 60.8%; Score 1889.8; DB 21; Length 3149;					
XX	XX	Best Local Similarity 81.4%; Pred. No. 0;					
XX	XX	Matches 2222; Conservative 0; Mismatches 504; Indels 5; Gaps 3;					
QY	50	GACCGCAAAACATGGTGTAGCAATCGGGCGCGAAGCGGAGCGGCGTCTCAGAGCTT	109				
DB	23	GGCTTGGGGACCATGGGAGGCGATCGGCGCGCAAGGGCGGAGGGGCGCCGAAGGACTT	82				
QY	110	CGGCGGGGACTCAAGTACAACTCCCGGTAGACACATGATGCTTGTAGAGGGTGT	169				
DB	83	CGGCGGGGACTCAAGTACAACTCCCGGCACGAGAAAGTGAATGGCTTGGAGGAAGGCT	142				
QY	170	GGAGTTCTCTGCGTCCGAAACAAATGCCAAGAAAGTGGAGAGCGAGGCGCCAGCGCTGGGT	229				
DB	143	GGAGTTCTCTGCGTCCGAAACAACTGCCAAGAAAGTGGAGAGCATGGCCCGGGGCGTGGGT	202				

Db 1283 CGTGGTCAACAGCAACAGCAACAGATCACAGTTGCTTCCACTCAGATCAGTCTCAAC 1342
Qy GGACACGGGTTCTAGTGTAGTACTCTCTAGACTCCAAAGACCGTGCCTCCAGGGAT 1429
Db 1343 CGACACGGCTTCTAGTGTAGTACTCTCTAGACTCCAAAGACCGTGCCTCCAGGGAT 1402
Qy 1430 GTTCATGTGCAAGACTGACAGGTGATCCGAAAGAACTGCGCTGCGACGGCTGCGGAGA 1489
Db 1403 GTTCATGTGCGCAACGGGGCGGTGTATCCGAAAGAGCTGCGCTGTGATGCTGCGGCGA 1462
Qy 1490 CTGCCCCGATTAATAGTATAGAGGTACTGCGGATGCAATGCCACCCACAGATTCAGGTG 1549
Db 1463 CTGACCGACACAGCGATGAGCTCAACTGCAAGTTGGACGCCCGCCACCAATTCAGGTG 1522
Qy 1550 CAAAAACCAAGTTCTGCAAGGCCCTCTCTGCGGTCTGTGACAGTGTCAACGACTGTGGGGA 1609
Db 1523 CAAGACAAAGTTCTGCAAGGCCCTCTCTGCGGTCTGTGACAGTGTGAACGACTGTGGGAGA 1582
Qy 1610 CGGAAGTGACAGGAGGCTGCAAGTGTCTGCTGGAGTTTCAAGTGTTCCTCAATGGGAA 1669
Db 1583 CAACAGCGACGAGCGGCTGCAAGTGTCTGCGGCCACAGACCTTCAGGTGTTCCTCAATGGGAA 1642
Qy 1670 GTGCTCTCCTCAGACCGACAGATGTATGGAAGGACAACTGTGGAGATGGGTCTGACGA 1729
Db 1643 GTGCTCTCGAAAGCCAGCAGTGCATGGAAGGACGACTGTGGGACGGGTCCGACGA 1702
Qy 1730 GGCTTCATGTGACAGCGTGAATGCTCTCTGTCACCAAAATATACCTACCGCTCCCAAAA 1789
Db 1703 GGCTCTCTGCCCAAGGTGAAGCTGTCTGTTGACCAACACACACCTACCGCTCCCTCAA 1762
Qy 1790 TGGCTCTGTCTGAGAGGCAACCCCTGAGTGTGATGGGAAGACGGAATGTAGCGATGG 1849
Db 1763 TGGCTCTGTCTGAGAGGCAACCCCTGAGTGTGACGGGAAGGAGACTGTAGCGACGG 1822
Qy 1850 CTCGATGAGAAACCTGCTGAGTGTGGGCTGCGATCTCTTACCAACAGAGTCCGCTGGT 1909
Db 1823 CTCAGATGAGAAAGCTGCGACTGTGGGCTGCGGTCAATTCACGAGACAGGCTGTGTGT 1882
Qy 1910 TGTGGCACGAATCGGACAGGCGAGTGGCTCTGCGAGTGAAGCTCCACAGCCCTGGG 1969
Db 1883 TGGGGGACCGATGCGATGAGGCGAGTGGCTCTGCGAGTGAAGCTGTAGCTCTGGG 1942
Qy 1970 CAGGGGCACTGTGTGGGCGCTGCTCACTCTCTGCTGAGTGGCTGTCTGAGCTCA 2029
Db 1943 CCAGGGCCACATCTGCGGTGCTTCCCTCACTCTCTCCCACTGGCTGTCTCTGCGCACA 2002
Qy 2030 TTGCTTTTCAGGATGACAAATTTCAAGTACTCAGACTACAGATGTGGAGCGCTTCCT 2089
Db 2003 CTGCTACATCATGACAGAGGATTCAGGTACTCAGACCCCAAGTGGAGCGCTTCCT 2062
Qy 2090 GGCTCTCTGACAGCAAGCGCAGTGGCTCTGCGGTGCGAGGCTGAAGCTCAAAACG 2149
Db 2063 GGGCTTGACAGCAAGCGCAGCGCCCTGGGTGCGAGGCGCAGGCTCAAGCG 2122
Qy 2150 TATCATCAACCACTCTCTCAATGATTTCACTTCACTATGACATCGCTGTGGA 2209
Db 2123 CATCATCTCCCACTCTCTCAATGATTTCACTTCACTATGACATCGCTGTGGA 2182
Qy 2210 GCTGGAGAGTCTGCTGAGTACAGCAGCTGCTGCGGCCCTATCTGCTGTGATGTAC 2269
Db 2183 GCTGGAGAAACCGCAGAGTACAGTCCATGCTGGTGGGCCCTATCTGCTGCGGACGCTC 2242
Qy 2270 CCATGTCTTCTGCTGGCAGGCTCTGGGTGACAGGCTGGGGGACACAAAGAGGG 2329
Db 2243 CCATGTCTTCTGCTGGCAGGCTCTGGGTGACAGGCTGGGGGACACCCAGATAGG 2302
Qy 2330 AGGTACCGGAGCGCTGATCTGCAAGAGGCTGATCCGTGTCTATCAACAGACCACTG 2389
Db 2303 AGGCACCTGGGCGCTGATCTCTGCAAAAGGCTGATCCGCGTCTATCAACAGACCACTG 2362
Qy 2390 TGAGGACCTCATGCGCAGCAGATCAACCCACGATGATGTGTGGGTTCCTCAGTGG 2449

Db 2363 CGAAGACCTCTGCGCAGCAGATCAAGCGCGCATGATGTGCTTCTCCTCAGCGG 2422
Qy 2450 GGGTGTGGACTCTCTGCGCAGGCTGACTCTGTGGGCCCTTGTCAAGCGCGGAGAAAGATGG 2509
Db 2423 CGGGTGTGGACTCTCTGCGCAGGCTGATTCGGGGGAGCCCTGTCTCCAGCGTGGAGCGGATGG 2482
Qy 2510 GCGAATCTTCCAGGCTGCTGTGCTGAGTGGGGTGAAGGCTGCGCTCAGAGGAAACAAGCC 2569
Db 2483 GCGGATCTTCCAGGCGGCTGTGCTGAGTGGGAGACGGCTGCTCAGAGAACAGCC 2542
Qy 2570 AGGCTGTACAAAGGCTCTCTGTAGTTCGGGACTGGATCAAGAGACACTGGGGTATA 2629
Db 2543 AGGCTGTACAAAGGCTCTCTGTGTTTCGGGACTGGATCAAGAGAACACTGGGGTATA 2602
Qy 2630 GCAGATGACAGACAGCGCAGCACCAACACCCACAGGGATGCCGATGCACACCTGG 2689
Db 2603 GGGGCGGG---GCCACCCAAATGTGTACCTGTGGGGGACCCATCTGCTCCACCCCACT 2659
Qy 2690 ATACAGGAGGAACTGACGACATTTATGCTGTGGCTTCCCCCCCCCAACCAACCA 2749
Db 2660 GTGCAGC-CCTGCGAGCTGGAGACTGGACCGCTGACTGCACAGCGGCC-CCAGAACATA 2717
Qy 2750 GACTGTGAATGCACTCTTACGACTCAGAT 2780
Db 2718 CACTGTGAATCAATCTCCAGGGCTCCAAAT 2748

RESULT 3

AAF28099 ID AAF28099 standard; cDNA; 3142 BP.
XX AAF28099; AC
XX AC
DT 06-JUN-2001 (first entry)
XX
DE Human membrane-type serine protease MT-SPI coding sequence.
XX
KW Human; membrane-type serine protease; MT-SPI; cancer; ss.
XX Homo sapiens.
OS
XX
FH Location/Qualifiers
FT CDS 37..2602
FT FT /*tag= a
FT FT /product= "MT-SPI"
XX
PN WO200123524-A2.
XX
PD 05-APR-2001.
XX
PF 02-OCT-2000; 2000WO-US27250.
XX
PR 30-SEP-1999; 99US-0410362.
XX
PA (REGC) UNIV CALIFORNIA.
XX
PI Craik CS, Takeuchi T, Shuman M;
XX
DR WPI; 2001-245002/25.
DR P-PSDB; AAB35465.
XX
PT New nucleic acid encoding a membrane type serine protease, useful for
PT the diagnosis, prognosis and treatment of cancer, particularly
PT metastatic cancers -
XX
PS Claim 4; Fig 1; 102pp; English.

The present invention provides the protein and coding sequences for the novel human membrane-type serine protease MT-SPI. Increased expression of this protein is associated with cancer, and so the sequences can be used in cancer diagnosis and the identification of treatments. The present sequence is the MT-SPI coding sequence.

SQ	Sequence 3142 BP; 635 A; 958 C; 965 G; 584 T; 0 other;	
	Query Match 60.7%; Score 1884.8; DB 22; Length 3142;	
	Best Local Similarity 82.8%; Pred. No. 0;	
	Matches 2164; Conservative 0; Mismatches 447; Indels 1; Gaps 1;	
QY	34 CTGAGACCGCGCATCGGACCGCCAAACACATGGGTAGCAATGGGGCCCGAAGCCCGAG 93	1054 TCTTCAGCTGCCAAGATGAGCAGCTGTGGCGCTTTTGTAGTGACACCCAGGGACAT 1113
DB	9 CTGAGACCGCGAGCGG-CCTCGGGAGACCATGGGGAGCGATCGGGCCCGCAAGGGCGAG 67	1028 TCTTCAGCTGCTAGGATGAGCAGCTGTGGAGCCGCTTTACGTAAGCCCGAGGGACAT 1087
QY	94 GGGGCTCTCAGGACTTCGGCGCGGACTCAAGTACAACTCCCGCTAGAGAAATGATG 153	1114 TTAGCAGCCCTACTATATCCAGGCCACTACCCGCCCAACATCAACTGCACATGGAATATCA 1173
DB	68 GGGGCCCCGAAGGACTTCGGCGCGGACTCAAGTACAACTCCCGGACGAGAAAGTGAATG 127	1088 TCAACAGCCCTACTATACCCAGGCCACTACCCACCAACATGACTGCACATGGAACATTG 1147
QY	154 GCTTTGAGGAGGCTGTGGAGTCTCTGCTCGCAACAAATGCAAGAAAGTGGAGAAGCGAG 213	1174 AGGTGCCCAACACCGGAACGTGAAGTGGCTTCAAACTCTTCTATCTGTGGTGGACCCCA 1233
DB	128 GCTTGAGGAGGCTGTGGAGTCTCTGCTCGCAACAAATGCAAGAAAGTGGAGAAGCGAG 187	1148 AGGTGCCCAACACCGAGCATGTGAAGTGGCTTCAAACTCTTCTATCTGTGGTGGACCCG 1207
QY	214 GCCCAGGCGCTGGGTGGTCTGGTGGCAGTGTGTTTCACTTCCCTTGTCTCTCCCTCA 273	1234 ACCTACAGTGGGCTCTCTGCACCAAGCACTATGTGAGATCAACGGGGAGAACTACTGCG 1293
DB	188 GCCCGGGGCGCTGGGTGGTCTGGCAGCCGTGTGATCGGCCCTCTCTTGGTCTTGCTGG 247	1208 GCGTGCTTCGGGGCAGCTGCCCCNAGGACTAAGTGGAGATCAATGGGGAGAACTACTGCG 1267
QY	274 TGGCTGGCTTGTGGTGGCACTTCCATTATCGGATGTGGGTTCAAAAAGTCTTCA 333	1294 GTGAGAGTCCCAAGTTTGTGGTGGAGCAGCAACAGCAGCAAGATTAAGTCCACTTCCATT 1353
DB	248 GGATCGGCTTTCCTGGTGGCACTTTCAGTACCGGAGCGTGGGTGCCAGAGGTTCTCA 307	1268 GAGAGAGTCCCAAGTTTCGTCTGTCACCGCAACAGCAACAGATCAAGATTCAGATTCGCTTCCACT 1327
QY	334 ATGGCCATCTGAGGATCACAAATGAGATCTTCTGATGCGTATGAGAACTCCACCTCCA 393	1354 CTGATCACTCGTACACGAGACACCGGTTCTTAGCTGAGTACCTCTCTACGACTCCAAAG 1413
DB	308 ATGGTACATGAGGATCACAAATGAGAAATTTTGGATGCTCTACGAGAACTCCCACTCCA 367	1328 CAGATCAGTCTTACACCGACACCGGCTTCTTAGCTGAATACCTCTCTACGACTCCAGTG 1387
QY	394 CAGAGTTTATCAGCTGGCGCAGCAGGAGTGAAGAGGCGCTGAAGTGTCTGTAAATGAAG 453	1414 ACCGTGCCCCAGGAGTGTTCATGTGCAAGACTGGACGCTGCATCCGAAAGGAACCTCGCT 1473
DB	368 CTGAGTTTGTAAAGCTGGCGCAGCAAGGTGAAGACGCGCTGAAGTGTCTGTACAGCGGAG 427	1388 ACCATGCGCGGGGCAAGTTTCACTGCGGCACCGGGCGGTGTATCCGGAAGGAGCTCGCT 1447
QY	454 TCCTGTCTGGTTCCTACCAAGAGTCCGCTGTAACTGCTTCAAGTGGGAGGAGTG 513	1474 GCGACGGCTGGGAGACTGCGCGGATTAAGTGTAGGAGTGTACTGCCGATGCAATGCCA 1533
DB	428 TCCATTCCTGGGCCCCCTACCAAGAGTCCGCTGTGACGGCTTTCAGGAGGAGGAGCG 487	1448 GTGATGGCTGGGCGGACTGACCGACCAAGGATGAGTCAACTGAGTTGCGACGCG 1507
QY	514 TCATCGCTACTACTGTGTAGAGTTCAGACTCCCGCCACACCTGGCAGAAAGGTTGATC 573	1534 CCACACAGTTACGTGCAAAAACAGTCTCTCAAGCCCTCTTCTGGGCTGTGACAGTG 1593
DB	488 TCATCGCTACTACTGTGTAGTTCAGACTCCCGCAGCACCTGGTGGAGAGGCGCGAGC 547	1508 GCGACAGTTTACGTGCAAGAACAGTCTCTCAAGCCCTCTTCTGGGCTGTGCGACAGTG 1567
QY	574 GCGCATGGCTGTGGAGGAGTGTAAATTCGACCCCGCAGCAGCGGCACTGAATTCCT 633	1594 TCACAGCTGTGGGAGCGGAAGTGAAGAGGCGCTGCAGCTGTCTCTGGGAGTTTCA 1653
DB	548 GCGTCATGGCGGAGGAGCGGTGTATGCTGTGCCCCCGCGCGCGCTCCCTGAAGTCTCT 607	1568 TGAAACGACTGGGAGACAACAGCAGCAGAGCGGGTGCAGTTGTCTCGGCCACAGCTTCA 1627
QY	634 TCGTGTAACTGTGTGGTGGCTTCCCATTTGACCCCAAGATGCTGCAGAGACTCAGG 693	1654 AGTGTTCGAATGGGAAGTGTCTCCCTCAGAGCCAGAAAGTGTAAATGGGAAGGACAATGTG 1713
DB	608 TTGTGGTCACTCAGTGTGGCTTTCCCAACGACTTCCAAAACAGTACAGAGACCCAGG 667	1628 GGTGTTCGAATGGGAAGTGTCTCTCGAAAAGCCAGCAGTGCATGGGAAGGAGCACTGTG 1687
QY	694 ACAACAGCTCAGTTTTCCTGCATGCCCATGGTGCAGCAGTGACACGTTCACTACCC 753	1714 GAGATGGGCTGCACAGAGCTTCATGTGACAGCGTGAATGTGCTCTTGCACCAATATA 1773
DB	668 ACAACAGCTCAGTTTTCCTGCATGCCCATGGTGCAGCAGTGATGCGCTTCAACACGC 727	1688 GGGACGGGTCGACAGAGGCTCTCTGCCCAAGGTGAACGTCTGTACCAACACA 1747
QY	754 CTGCTTCCCAACAGTTCCTACCGCGGATCCCGCTGCCAGTGGTCTCTGGGGGG 813	1774 CTTACCGCTGCCAAATATGGCTCTGTCTGAGCAAGGGCAACCTGTAGTGTGTGGGAAGA 1833
DB	728 CCGCTTCCCTGACAGCCCTACCCCGCTCATGCCCGCTGACAGTGGGCGCTCGCGGGGG 787	1748 CTTACCGCTGCTCAATGGGCTCTGTCTGAGCAAGGGCAACCTGTAGTGTGACGGGAAG 1807
QY	814 ACGCCGACTCTGTGCTGAGCTCACCTTCCGAAAGCTTTGATGTGCTCTCTGTGATGAGC 873	1834 CGGACTGTAGCGATGCTCGATGAGAAAACCTGTGACTGTGGGCTGCATCTTTACCA 1893
DB	788 ACGCCGACTCAGTGTGAGCTCACCTTCCGAGCTTTGACCTTGTGCTGTGAGCAGC 847	1808 AGGACTGTAGCGACGCTCAGATGAGAGGACTGTGCGCTGTGGGCTGCCTCACTCAGCA 1867
QY	874 ATGGCAGTGAAGTGTCAACCGTGTATGATGCTGTAGCGCCCACTGGAACCCCAACGCTGTGG 933	1894 AACAGGCTCGCGTGGTGGTGGCAGCAATTCGCGACGAGGCGAGTGGCGCTTGGCAGGTGA 1953
DB	848 GCGGACGAGCCTGGTGAACGCTGTACAAACCCCTGAGCCCTGAGCCCTGAGCCCTGAG 907	1868 GACAGGCTGTGTGTGGGGGCAAGATGCGGATGAGGCGGAGTGGCGCTTGGCAGGTAA 1927
QY	934 TGGGCTGTGGCAGCTTCTCACCTCTTACACCTGATCTTCTCTCTCTCTCTCTCTCTCT 993	1954 GCGTCCACGCGCTTGGGCGCAGGGGCACTTGTGTGGGCGCTCGCTCATCTCTCTCTCTCT 2013
DB	908 TGCAGTTGTGTGGCAGCTTCT 967	1928 GCGTGCATCTCTGGGCGCAGGGGCACTCTGGGCTGTCTTCTCTCTCTCTCTCTCTCTCTCT 1987
QY	994 TCTTCT 1053	2014 TGGTCT 2073
DB	968 TCTTCT 1027	1988 TGGTCT 2047
		2074 TGTGACGCGCTTCTCTGGGCTCTCTGGGACAGAGCAAGCGGAGTGGCTCTCTGGGCTCAGG 2133
		2048 AGTGGACGCGCTTCTCTGGGCTTGCACGACACGAGCCAGCGCGCCCTCTGGGCTCAGG 2107

2134 AGCTAGAGCTCAACAGPATCATACCCACACCTCTCTCAATGATTTACCTTGGACTATG 2193
 2108 AGCGAGGCTCAAGCGCATCATCTCCACACCTCTCTCAATGATTTACCTTGGACTATG 2167
 2194 ACATCGCTTCTGGAGCTGAGAGTCTGGTGGAGTACAGACCGTCTGGTGGCCCATCT 2253
 2168 ACATCGGCTCTGGAGCTGAGAGTCTGGTGGAGTACAGACCGTCTGGTGGCCCATCT 2227
 2254 GCCTGCTGATGTACCATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2313
 2228 GCCTGCTGAGCT 2287
 2314 GGCACACAAAGAGGAGGAGTACCGAGCTGTGATCTCTCTCTCTCTCTCTCTCTCTCT 2373
 2288 GACACACCCAGTATGGAGCACTCTGGCGCTGTATCTCTCTCTCTCTCTCTCTCTCTCT 2347
 2374 TCACACAGACCACTGTGAGGACCTCTATGCGGAGCAGATCACCCACGAAATGATGTG 2433
 2348 TCACACAGACCACTGTGAGGACCTCTATGCGGAGCAGATCACCCACGAAATGATGTG 2407
 2434 TGGGTTTCTCTCAGTGGGGGTGTGACTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2493
 2408 TGGGTTTCTCTCAGTGGGGGTGTGACTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2467
 2494 GCGCGAGAAAGATGGCGCAATGTTCCAGGCTGGTGTGGTGGAGTGGGGTCAAGGCTGG 2553
 2468 GCGTGGAGGCGGATGGCGGATCTTCCAGGCTGGTGTGGTGGAGTGGGGGAGCGGCTGG 2527
 2554 CTGAGGAGACACCGAGGCTGTACACAGGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2613
 2528 CTGAGGAGACACCGAGGCTGTACACAGGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2587
 2614 AGCACACTGGGGTATAGCAGCATGGACAGACA 2645
 2588 AGAACCTGGGTATAGGGCGGGGACACCA 2619

RESULT 4

AAX87815
 ID AAX87815 standard; cDNA; 3147 BP.
 XX
 AC AAX87815;
 XX
 DT 09-NOV-1999 (first entry)
 XX
 DE Tumour antigen derived gene-15 (TADG-15) cDNA.
 XX
 KW Tumour antigen derived gene-15; TADG-15; serine protease; human;
 KW breast cancer; ovary cancer; carcinoma; diagnosis; ss.
 XX
 OS Homo sapiens.
 XX
 PH Key Location/Qualifiers
 FT CDS 23..2590
 FT /*tag= a
 XX
 WO9942120-A1.
 XX
 PD 26-AUG-1999.
 XX
 PF 18-FEB-1999; 99WO-US03436.
 XX
 PR 20-FEB-1998; 98US-0027337.
 XX
 PA (UYAR-) UNIV ARKANSAS.
 XX
 PI O'Brien TJ, Tanimoto H;
 XX WPI; 1999-527418/44.
 DR P-PSDB; AAY06671.
 XX
 PT A new extracellular serine protease for diagnosis of neoplastic disease

XX
 PS
 XX Claim 2; Fig 9; 71pp; English.
 CC This is the nucleotide sequence of an isolated cDNA that codes for
 CC an extracellular serine protease, termed tumour antigen derived
 CC gene-15 protein (see AAY06671), that is overexpressed in breast and
 CC ovarian carcinomas. The TADG-15 gene can be used as a diagnostic
 CC and therapeutic target in ovarian carcinoma and other carcinomas
 CC including breast, prostate, lung and colon. The TADG-15 cDNA was
 CC isolated from ovarian carcinoma by PCR using primers directed to
 CC conserved areas of the serine protease family. The invention
 CC also provides: a vector that is capable of expressing DNA encoding
 CC TADG-15 protein; host cells selected from bacterial cells
 CC (especially *Escherichia coli*), mammalian cells, plant cells and
 CC insect cells; and a method of detecting expression of TADG-15
 CC protein using a hybridisation probe.
 XX
 SQ Sequence 3147 BP; 654 A; 952 C; 958 G; 583 T; 0 other;
 Query Match 60.6%; Score 1883.2; DB 20; Length 3147;
 Best Local Similarity 81.2%; Pred. No. 0;
 Matches 2223; Conservative 0; Mismatches 508; Indels 5; Gaps 3;
 QY 45 GATCGGACCGCCAAAACCAATGGGTAGCAATCGGGCGCGCAAGCCGAGGGGCTCTCAG 104
 Db 5 GAGCGGCTCGGGGTACCATGGGAGCGATCGGGCGCGCAAGCGGAGGGGCCGAAG 64
 QY 105 GACTTCGGCGGGGACTCAAGTACAACTCCCGCTAGAGACATGATGCTTTGAGGAG 164
 Db 65 GACTTCGGCGGGGACTCAAGTACAACTCCCGCTAGAGACATGATGCTTTGAGGAA 124
 QY 165 GGTGTGGAGTTCCTGCTGCGAACTATGCGAAAGTGGAGAGCGAGGCCCGCAGCGC 224
 Db 125 GCGGTGGAGTTCCTGCGAGTCAACACGTCAAGAGGTGGAAGCATGCGCGGGCGC 184
 QY 225 TGGGTGGTGTGGGAGTCTGTTCAGTTCCTCTTGTCTCTCTCTCTCTCTCTCTCTCT 284
 Db 185 TGGGTGGTGTGGGAGTCTGTTCAGTTCCTCTTGTCTCTCTCTCTCTCTCTCTCTCT 244
 QY 285 CTGGTGGGCACTTCCATATCGGAATGTGGGGTTCAAAAGTCTTCAATGCCATCTG 344
 Db 245 CTGGTGGGCACTTCCATATCGGAATGTGGGGTTCAAAAGTCTTCAATGCCATCTG 304
 QY 345 AGGATCACAAAATGAGATCTTCTGGATGCGTATGAGAACTCCACCTCCACAGAGTTATC 404
 Db 305 AGGATCACAAAATGAGATCTTCTGGATGCGTATGAGAACTCCACCTCCACAGTTGA 364
 QY 405 AGCTGCGCAGCCAGGTGAGAGGCGCTGAAAGTCTGTATCAATGAAGTCCCTGCTCTG 464
 Db 365 AGCTGCGCAGCCAGGTGAGAGGCGCTGAAAGTCTGTATCAATGAAGTCCCTGCTCTG 424
 QY 465 GGTCCCTTACCACAAAGATCGGCTGTAACTGCTTTCAGTGGAGGCGAGTGTATCGCCTAC 524
 Db 425 GGCCTCTTACCACAAAGATCGGCTGTAACTGCTTTCAGTGGAGGCGAGTGTATCGCCTAC 484
 QY 525 TACTGTTCAAGTTCAGCATTCCTCCCACTCCGACAGAGAGGTGTATCGCGCATGGCT 584
 Db 485 TACTGTTCAAGTTCAGCATTCCTCCCACTCCGACAGAGAGGTGTATCGCGCATGGCT 544
 QY 585 GTGAGGAGTGTGAACATTCCTCCCACTCCGACAGAGAGGTGTATCGCGCATGGCT 644
 Db 545 GAGGAGGAGTGTGAACATTCCTCCCACTCCGACAGAGAGGTGTATCGCGCATGGCT 604
 QY 645 TCTGTGTGGCTTCTCCCACTTCGACCTCCGACAGAGAGTGTATCGCGCATGGCT 704
 Db 605 TCAAGTGTGGCTTCTCCCACTTCGACCTCCGACAGAGAGTGTATCGCGCATGGCT 664
 QY 705 AGTTTGGCTGTGATGCCCATGGTGTGAGCAGTGTATCTACTACCTGCTTCCCTCC 764
 Db 665 AGTTTGGCTGTGATGCCCATGGTGTGAGCAGTGTATCTACTACCTGCTTCCCTCC 724
 QY 765 AACAGTCCCTTACCCGCGGATGCTCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 824

MTSP; protease domain; neoplastic disease; tumour; cancer; cytostatic; lung carcinoma; colon adenocarcinoma; ovarian carcinoma; gene therapy; matriptase; ds.

XX Homo sapiens.

OS
XX Location/Qualifiers
FH Key 23..2590
CD CDS

FT /*tag= a
FT /product= "Human transmembrane serine protease (MTSP) 1"

XX WO200157194-A2.

XX 09-AUG-2001.

XX 02-FEB-2001; 2001WO-US03471.

XX 03-FEB-2000; 2000US-0179982.

XX 18-FEB-2000; 2000US-0183542.

XX 22-JUN-2000; 2000US-0213124.

XX 26-JUL-2000; 2000US-0220970.

XX 08-SEP-2000; 2000US-0657986.

XX 22-SEP-2000; 2000US-0234840.

XX (CORV-) CORVAS INT INC.

XX Madison EL, Ong EO, Yeh J;

XX WPI; 2001-488877/53.

XX P-PSDB; AAE06930.

XX Novel single chain polypeptide comprising protease domain of type-II membrane-type serine protease or its catalytically active portion useful for treating and preventing cancer and tumor -

XX Claim 14; Page 191-195; 256pp; English.

XX The invention relates to transmembrane serine proteases and their corresponding nucleotides and the protease domain of a type-II membrane-type serine protease (MTSP). MTSP is useful for identifying compounds that modulate or inhibit its proteolytic activity and for formulating a medicament for treating neoplastic disease. MTSP and its corresponding nucleotides are useful in preventing or treating tumours or cancers such as lung carcinoma, colon adenocarcinoma and ovarian carcinoma, in diagnostics and in hybridisation assays. MTSP is useful as a diagnostic marker for tumour development, growth and/or progression and as immunogens to generate antibodies that specifically bind to it. MTSP DNA is useful in a yeast two-hybrid system and in gene therapy. The present sequence is a DNA encoding human MTSP1 protein (also called matriptase).

XX Sequence 3147 BP; 654 A; 952 C; 958 G; 583 T; 0 other;

Query Match 60.6%; Score 1883.2; DB 22; Length 3147;
Best Local Similarity 81.2%; Pred. No. 0;
Matches 2223; Conservative 0; Mismatches 508; Indels 5; Gaps 3;

QY 45 GATCGGACGGCAAAACCATGGGTAGCAATCGGGCGCGCAAGGCGGGGGGCTCTCAG 104
DB 5 GAGCGGCTCGGGGTACCATGGGAGCGATCGGGCCCGCAAGGCGGGGGCCGAAG 64

QY 105 GACTTCGGCGGGGACTCAAGTCAAACTCCCGGCTAGAGAACATGAATGGCTTTGAGGAG 164
DB 65 GACTTCGGCGGGGACTCAAGTCAAACTCCCGGCGACGAGAAAGTGAATGGCTTGGAGGAA 124

QY 165 GGTGTGGATTCCTGCTCGGCAAAATGCGAAGAGTGGAGAGCGAGGCCCGCGCGC 224
DB 125 GGCCTGGAGTTCTTGCCAGTCAACACGTCGAAGAGGTGGAAGATGGCCCGGCGC 184

QY 225 TGGGTGGTGGTGGGAGTCTGTTCAGCTCTCTTGGCTCTCCCTCATGGCTGGCTTG 284
DB 185 TGGGTGGTGGTGGGAGCGGCTGATCGGCTCTCTTGGCTCTCTTGGGATCGGCTTC 244

QY 285 CTGGTGGGCACTTCATTATCGAATGTGCGGGTTCAAAAAGTCTTCAATGGCCATCTG 344
DB 245 CTGGTGGGCACTTCATTATCGAATGTGCGGGTTCAAAAAGTCTTCAATGGCTACATG 304
QY 345 AGGATCAAAATGAGATCTTTCTGGATGCGTATGAGAACTCCACTCCACAGAGTTTATC 404
DB 305 AGGATCAAAATGAGATCTTTCTGGATGCGTATGAGAACTCCACTCCACTGAGTTTGTG 364
QY 405 AGCTGCCGACGAGTGAAGGAGGCGCTGAAGTCTGTACATGAGTCCCTGCTCTG 464
DB 365 AGCTGCCGACGAGTGAAGGAGGCGCTGAAGTCTGTACATGAGTCCCTGCTCTG 424
QY 465 GGTCTCTTACCAAGAGTGGCTGTAATGCGCTTCACTGAGGCGAGTGTATCGCTAC 524
DB 425 GGTCTCTTACCAAGAGTGGCTGTAATGCGCTTCACTGAGGCGAGTGTATCGCTAC 484
QY 525 TACTGTCTCAGATTCAGCATCCCGCCCACTGTCGAGAGAGTGTATCGCGCATGGCT 584
DB 485 TACTGTCTCAGATTCAGCATCCCGCCCACTGTCGAGAGAGTGTATCGCGCATGGCT 544
QY 585 GTGGAGCGAGTGTGAACATTTGCCACCGGAGCAGCGGCACTGTAATCTTCTGCTCAACA 644
DB 545 GAGGAGCGAGTGTGAACATTTGCCACCGGAGCAGCGGCACTGTAATCTTCTGCTCAACC 604
QY 645 TCTGTGTGGTGGCTTCCCGCATTTGACCCCGAGAGTGTGCGAGAGGACTCAGAGCAACAGCTGC 704
DB 605 TCAGTGGTGGCTTCCCGCATTTGACCCCGAGAGTGTGCGAGAGGACTCAGAGCAACAGCTGC 664
QY 705 AGTTTGGCTGTGATGCGCATGTCGAGAGTGTGACGCTTCACTACCTCGGCTTCCCT 764
DB 665 AGCTTTGGCTGTGACGCGCGGCTGTGAGCTGATGCGCTTCACTACGCGCGGCTTCCCT 724
QY 765 AACAGTCTCTTACCGGCGCATGCGCTGCGAGTGGTCTGCGGGGAGCGGAGCTCT 824
DB 725 GACAGCCCTTACCGGCGCATGCGCTGCGAGTGGGCTTGGCGGGGAGCGGAGCTCA 784
QY 825 GTGCTGAGCTCACCTTCCGAAGCTTTGATGTGCTCTCTGTGATGAGCATGGCAGTGC 884
DB 785 GTGCTGAGCTCACCTTCCCGAGCTTTGACCTTGGCTTGGCGAGCGCGGAGCGAG 844
QY 885 CTGGTCACTGTATGATGAGCTTGAAGCCCATGGAACCCCACTGCTGTGCTGGCTGTGT 944
DB 845 CTGGTCACTGTATGATGAGCTTGAAGCCCATGGAACCCCACTGCTGTGCTGGCTGTGT 904
QY 945 GGCACTTCTCACCTCTCAACCTGACCTTCTCTCTCTCCCAAGACGCTTCTCTGTC 1004
DB 905 GGCACTTCTCACCTCTCAACCTGACCTTCTCTCTCTCCCAAGACGCTTCTCTGTC 964
QY 1005 ACCTGTATACCAATATCTGACCGCGCATCTCTGGCTTTGAGGCGCACTTTCTTCCAGCTG 1064
DB 965 ACCTGTATACCAATATCTGACCGCGCATCTCTGGCTTTGAGGCGCACTTTCTTCCAGCTG 1024
QY 1065 CCCAAGTGAAGAGCTGTGGGGCTTTTGAAGTGAACCCCAAGGAGCATTTAGCGAGCCC 1124
DB 1025 CCTAGGATGAGAGCTGTGGAGGCGGCTTACGTAAGCCCGAGGAGCATTTCAACAGGCCCC 1084
QY 1125 TACTATCCAGGCGCACTACCGCGCCCAACATCAACTGACATGGAATATCAAGGTGCCCAAC 1184
DB 1085 TACTATCCAGGCGCACTACCGCGCCCAACATTTGACATGGAATGGAATGAGGTGCCCAAC 1144
QY 1185 AACCGGAAGTGAAGTGGCTTCAAACTCTTATCTGTGTGAGACCCCAACGATACAGTGT 1244
DB 1145 AACCGGAAGTGAAGTGGCTTCAAACTCTTATCTGTGTGAGACCCCAACGATACAGTGT 1204
QY 1245 GGCTCTGACCAAGAGTGTGAGATCAACGGGAGAGAGTACTGCGGTGAGAGTGC 1304
DB 1205 GGCACTTGGCCCCAAGGAGTGTGAGATCAATGGGAGAGAAATACTGCGGAGAGAGTGC 1264
QY 1305 CAGTTTGTGTGAGAGCAACAGCAGCAAGTATACAGTCCACTTCCACTTCTGATCACTCG 1364
DB 1265 CAGTTTGTGTGAGAGCAACAGCAGCAAGTATACAGTCCACTTCCACTTCTGATCACTCG 1324
QY 1365 TACACGGAGACCGGGGTTCTTCTAGTGTAGTACCTCTCTTACGACTCCAAAGACCGGTGCCCA 1424

1325 TACACGACACCGGCTTCTTAGCTGAATACCTCTCTACGACTCCAGTGCACCCATGCCCG 1384
1425 GGGATGTTCTATGTCGAAGACTGAGCGTGCATCGAAGGAACTGCGCTCGAGCGGTGG 1484
1385 GGGAGTTTCACTGCGCCACCGGGCGGTGTATCGGAGGAGCTGCGCTGTGATGGCTGG 1444
1485 GCAGACTGCCCGGATTTATGATGATGAGCGTTTACTGCCGATGCAATGCAACCCACCACTTC 1544
1445 GCCGACTGCACCGACACAGCGATGAGCTCAATGACAGTTGGACGCGCGCCACAGTTC 1504
1545 ACGTGCAAAACACAGTTCTGCAAGCCCTCTCTGCGGTCTGTGACAGTGTCAAGCACTGT 1604
1505 ACGTGCAAGAACAGTTCTGCAAGCCCTCTCTGCGGTCTGTGACAGTGTGAAACGACTGC 1564
1605 GGGACGCAAGTGTACGAGGAGGGCTGCAAGTGTCTCTGCTGGAGTTTCAAGTGTTCCAAT 1664
1565 GGAGACAACACGACGAGCGGGTGCAGTTGTCCGCGCCACGACCTTCAGTGTTCCAAT 1624
1665 GGGAAAGTGTCTCTCAGAGCGAGAGTAAATGGGAAGACAACTGTGTGAGATGGGTCT 1724
1625 GGGAAAGTGTCTCTCAGAAAGCGACAGTGCATGGGAAGACGACTGTGGGACGGGTCC 1684
1725 GACGAGCTTCATGTGACAGGTGAATGTCTCTCTGCTGCAACCAATATACCTACCGCTGC 1784
1685 GACGAGCTTCCTCCCAAGGTGAAGCTGTCTCTGCTGCAACCAATATACCTACCGCTGC 1744
1785 CAAATGCGCTCTCTGAGCAAGGCAACCTCTGAGTGTGATGGGAAGACGACTGTAGC 1844
1745 CTCAATGGCTCTCTGAGCAAGGCAACCTCTGAGTGTGACGGGAAGAGGACTGTAGC 1804
1845 GATGGCTCCGATGAGAAATGTGACTGTGGGTGCGATCTTTTACCAACAGGCTCGC 1904
1805 GACGGCTCAGATGAGAGGACTGCGACTGTGGGTGCGGTCTTACGAGACAGGCTCGT 1864
1905 GTGGTTGGTGGCAAGTGGAGGCGGAGTGGCGCTTGGCAGAGTGGCGCTTGGCAGTGGCT 1924
1865 GTGGTTGGGCGCAGTGGATGAGGGGAGTGGCGCTTGGCAGAGTGGCGCTTGGCAGTGGCT 1924
1965 CTGGGCGAGGCGCACTGTGTGGGGCGCTCGCTCATCTCTCTGACTGGCTGGTCTTGCA 2024
1925 CTGGGCGAGGCGCACATCTCGGGTCTCTCTCATCTCTCTCCCACTGGCTGGTCTTGCC 1984
2025 GCTCATGCTTTCAGATGAGAAATTTCAAGTATCTGACATACAGATGTGAGCGGCC 2084
1985 GCACACTGCTACATGATGATGACAGAGGATTCAGTACTCTCAGACCCACGAGTGGCGCC 2044
2085 TTCTGGGTCTGTGGACACAGAGCAAGCGAGTGCCTCTGCGGTGACAGGAGTGAAGCTC 2144
2045 TTCTGGGTCTGACGACACAGAGCGAGCGCCCTCTGGGTGACAGGAGCGAGGCTC 2104
2145 AAGCGTATCATACCCACCTCTCTCAATGATTTTCACTTCCGATATGACATCGCCTTG 2204
2105 AAGCGTATCATCTCCACCTCTCTCAATGATTTTCACTTCCGATATGACATCGCCTTG 2164
2205 CTGAGCTGAGAGTCTGGTGGAGTACAGCAGCGTCTGTCGCGCCCATCTGCTGCTGAT 2264
2165 CTGAGCTGAGAGAACCGGACAGATACAGTCTCATGTTGGCGCCCATCTGCTGCTGAT 2224
2265 GCTACCCATCTCTCTCTGTCGACAGGCGCATCTGGGTGACAGGCTGGGGGACACAAA 2324
2225 GCCTCCCATGCTCTCTCTGTCGCGAGGCGCATCTGGGTGACAGGCTGGGGACACCCAG 2284
2325 GAGGAGGTACCGAGCGCTGATCTCTGAGAGGTTGAGATCCGTGTCAATCAACAGAC 2384
2285 TATGAGGCGCTGGCGGCTGATCTCTCAAAAGGTTGAGATCCGCGTTCATCAACAGAC 2344
2385 ACCTGTGAGAGCTCATGCGCGACAGATCAACCCAGCAATGATGTGTGGTTCCTC 2444
2345 ACCTGTGAGAGCTCTCTGCGCGACAGATCACCGCGCATGATGTGTGGTTCCTC 2404
2445 AGTGGGGGTGTGACTCTCTGCGAGGTTGACTCTGTGGCGCCCTTGTCAAGCGCGAGAA 2504

2405 AGCGGGCGGTGACTCTCTCCAGGCTGATTTCCGGGGACCCCTCTCCAGGCTGGAGCG 2464
2505 GATGGCGGATGTTCCAGGCTGCTGTGTGCTGAGCTGGGTGAAGGCTGCGCTCAGAGGAC 2564
2465 GATGGCGGATGTTCCAGGCTGCTGTGTGCTGAGCTGGGTGAAGGCTGCGCTCAGAGGAC 2524
2565 AAGCCAGGCTGTACACAAGGCTCCCTCTGATTTCCGGGACTGGATCAAGAGCACTCTGGG 2624
2525 AAGCCAGGCTGTACACAAGGCTCCCTCTGATTTCCGGGACTGGATCAAGAGCACTCTGGG 2584
2625 GTATAGCAGATGAGCAGACAGCGGACCAACACACCAAGGATGCCGACATGACACA 2684
2585 GTATAGGCGCGGGG---GCCACCCAAATGTGTACACTGCGGGGCCACCCATCGTCCACC 2641
2685 CTGGGATACAGAGAGAGAAACACTGACGACATTTATGCTGTGGCTCCCGCCCAACACA 2744
2642 CCACTGTGACG---CTTGCGAGCTGGAGACTGGACCGCTGACTGACACAGCGGCC---CCAGA 2699
2745 ACCGAGCTGTGAACCTGCACTCTCTTAGGACTCAGAGT 2780
2700 ACATACACTGTGAACCTCAATCTCCAGGCTCCAAAT 2735

RESULT 6
AAD13155
ID AAD13155 standard; DNA; 3147 BP.
XX AAD13155;
XX AC
XX AC
DT 16-OCT-2001 (first entry)
XX Human membrane-type serine protease (MTSP) 1 protease domain DNA.
DE Human; transmembrane serine protease; membrane-type serine protease;
XX MTSP; protease domain; neoplastic disease; tumour; cancer; cytostatic;
KW lung carcinoma; colon adenocarcinoma; ovarian carcinoma; gene therapy;
KW matriptase; ds.
XX Homo sapiens.
XX Key
XX CDS
XX Location/Qualifiers
1865..2590
/*tag= a
/product= "Human transmembrane serine protease
/note= "CDS does not include start codon"
/partial
XX WC200157194-A2.
XX 09-AUG-2001.
XX 02-FEB-2001; 2001WO-US03471.
XX 03-FEB-2000; 2000US-0179982.
XX 18-FEB-2000; 2000US-0183542.
XX 22-JUN-2000; 2000US-0213124.
XX 26-JUL-2000; 2000US-0220970.
XX 08-SEP-2000; 2000US-0657986.
XX 22-SEP-2000; 2000US-0234840.
XX (CORV-) CORVAS INT INC.
XX Madison EL, Ong BO, Yeh J;
XX WPI; 2001-488877/53.
XX P-PSDB; AAE06936.
XX Novel single chain polypeptide comprising protease domain of type-II
PT membrane-type serine protease or its catalytically active portion
XX useful for treating and preventing cancer and tumor
XX Example 6; Page 225-227; 256pp; English.
PS

XX The invention relates to transmembrane serine proteases and their
CC corresponding nucleotides and the protease domain of a type-II
CC membrane-type serine protease (MTSP). MTSP is useful for identifying
CC compounds that modulate or inhibit its proteolytic activity and for
CC formulating a medicament for treating neoplastic disease. MTSP and
CC its corresponding nucleotides are useful in preventing or treating
CC tumours or cancers such as lung carcinoma, colon adenocarcinoma and
CC ovarian carcinoma, in diagnostics and in hybridisation assays. MTSP
CC is useful as a diagnostic marker for tumour development, growth and/or
CC progression and as immunogens to generate antibodies that specifically
CC bind to it. MTSP DNA is useful in a yeast two-hybrid system and in
CC gene therapy. The present sequence is a DNA encoding protease domain of
CC human MTSP1 protein (also called matritase).

XX
SQ Sequence 3147 BP; 654 A; 952 C; 958 G; 583 T; 0 other;

Query Match 60.6%; Score 1883.2; DB 22; Length 3147;
Best Local Similarity 81.2%; Pred. No. 0;
Matches 2223; Conservative 0; Mismatches 508; Indels 5; Gaps 3;

QY 45 GATCGGACCCCAAAACCATGGGTAGCAATCGGGCGCGCAAGCCGAGGGGGCTCTCAG 104
Db 5 GAGCGGCTTCGGGGTACATGGGAGGAGTCGGGCCCGCAAGGGCGGGGGCCCGAAG 64

QY 105 GACTTCGGCGCGGACTCAAGTCAAACTCCCGGCTAGAGAACATGAATGGCTTGGAGAG 164
Db 65 GACTTCGGCGCGGACTCAAGTCAAACTCCCGGCAAGAGAAAGTGAATGGCTTGGAGAA 124

QY 165 GGTGTGAGTTCCTGCTGCGAAACATGCGCAAGAAAGTGGAGAAAGCGGCCCGAGCGC 224
Db 125 GGGGTGAGTTCCTGCGGAGTCAACAGCGTCAAGAAAGGTGGAAAAGCATGGCCGGGCGC 184

QY 225 TGGGTGTGCTGGGCGGAGTGTCTGTCAGTCTCTCTGCTCTCTCTCTCTCTCTCTCT 284
Db 185 TGGGTGTGCTGGGCGGAGTGTCTGTCAGTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 244

QY 285 CTGGTGTGGCACTTCCATTAATCGGAATGTGGGGTTCAAAAGTCTTCAATGGCCATCTG 344
Db 245 CTGGTGTGGCACTTTCAGTACCGGAGCTGCTGCTCCAGAAAGTCTTCAATGGCTACATG 304

QY 345 AGGATCAAAATGAGATCTTCTGGATGCGGTATGAGAACTCCACCTCCACAGAGTTATC 404
Db 305 AGGATCAAAATGAGAAATTTGTGGATGCGCTTACAGAACTCCAACTCCACTGAGTTTGA 364

QY 405 AGCTGCGCAGCGAGGTGAAGGAGCGCTGAAGCTGCTGTACATGAATGAATCCCTGTCTG 464
Db 365 AGCTGCGCAGCGAGGTGAAGGAGCGCTGAGCTGTGTACGGAGTCCCATTCCTG 424

QY 465 GGTCCCTACCAAGAAAGTGGGTGTAACTGCTTCAAGTGGAGGAGTGTATCGCCCTAC 524
Db 425 GGCCTCTACCAAGAGAGTGGGTGTGACGGCTTTCAGCGAGGCGAGCGTCAATCGCTAC 484

QY 525 TACTGTTCAGAGTTACAGATCCCGCCACACCTGCGAGAGAGTTGATCGGCCATGGCT 584
Db 485 TACTGTGTCTGAGTTTACAGATCCCGCCAGCACCTGCTGGAGGAGGCGGAGCGTCAATG 544

QY 585 GTGAGCGAGTGTAAATGTCACCGGAGCGGAGCTGAAATCTCTGCTCTAACA 644
Db 545 GAGAGCGGTAGTATGCTGCTCCCGCGGGCGGCTCCCTGAAAGTCTCTTGTGGTCAACC 604

QY 645 TCTGTGTGGCTTCCCATATGACCCAGAAATGCTGAGAGGAGTCAAGGACCAACAGCTGC 704
Db 605 TCAGTGTGGTGTTCCTCCACCGGACTCCAAAACAGTACAGAGGAGCCAGGACCAACAGCTGC 664

QY 705 AGTTTTCCTGATGATCCCATGTTGTCAGAGTGTGACAGCTTCACTACCTCGGCTTCCC 764
Db 665 AGCTTTGGCTGTGACAGCGCCCGGCTGTGAGCTGTATGCGCTTTCACCAAGCGCGGCTTCC 724

QY 765 AACAGTCTCTACCGCGGAGTGTGCGGCTGCGAGTGGGTCTCTGCGGGGGAGCGGAGCTCT 824
Db 725 GACAGCCCTACCCGCTCTATGCGCGCTGCCAGTGGGCGCTGCGGGGGAGCGCCGACTCA 784

QY 825 GTGCTGAGCCTCACCTTTCGGAAGCTTTGATGTGCTGCTCCCTGTGATGAGCATGGCAGTGAC 884
Db 785 GTGCTGAGCCTCACCTTTCGGAAGCTTTGATGTGCTGCTCCCTGTGATGAGCATGGCAGTGAC 844

QY 885 CTGGTCAACCGTGTATGATAGCTGAGCCCATGGAACCCACACGCTGTGGTGGCGCTGTGT 944
Db 845 CTGGTCAACCGTGTATGATAGCTGAGCCCATGGAACCCACACGCTGTGGTGGCGCTGTGT 904

QY 945 GGCACCTTCTCAACCTCTCTACACCTGACCTTCTCTCTCTCCCGAGAACGCTTCTCTGTGTC 1004
Db 905 GGCACCTTCTCAACCT 964

QY 1005 ACCTGTATTAACCAATATGACCCGCGCACATCTCTGGCTTTTGGAGCCACTTTCTTCTCCAGCTG 1064
Db 965 ACCTGTATTAACCAATATGACCCGCGCACATCTCTGGCTTTTGGAGCCACTTTCTTCTCCAGCTG 1024

QY 1065 CCAAGATGAGCAGCTGTGGCGGCTTTTGTAGTGAACACCAAGGAGCAATTTAGCAGCCCC 1124
Db 1025 CTTAGGATGAGCAGCTGTGGAGGCGGCTTACGTAAAGCCCGAGGGAGCATTTCAACAGCCCC 1084

QY 1125 TACTATCCAGGCGCACTACCCGCGCAACATCACTGACATGCAATGCAATCAAGTGTGAGTGTG 1184
Db 1085 TACTATCCAGGCGCACTACCCGCGCAACATTTGACTGCAATGAGGAGTGTGAGGTTGCCAAC 1144

QY 1185 AACCGGAACCTGAGAGTGGCGCTTCAAACTCTTCTATCTGCTGGACCCCAACGCTACCACTG 1244
Db 1145 AACCGATGTGAGAGTGGCGCTTCAAACTCTTCTATCTGCTGGAGCCCGGCTGCTGCG 1204

QY 1245 GGTCTCTGACCAAGAGCTATGTGGAGATCAACGGGGAGAAAGTACTGCGGTGAGAGGTCC 1304
Db 1205 GGCACCTTCTGCGGAGGAGTACCTGAGATCAATGGGGAGAAATACTGCGGAGAGAGGTCC 1264

QY 1305 CAGTTTGTGTGAGGAGCAGCAGCAGCAAGATTTACAGTCCACTTCCATTTCTGATCTACTCG 1364
Db 1265 CAGTTTGTGTGAGGAGCAGCAGCAGCAAGATTTACAGTCCACTTCCATTTCTGATCTACTCG 1324

QY 1365 TACACGAGCACCGGGTTCTTAGCTGAGTACTCTCTCTACGACTCCCAACGAGCCGCTGCCCA 1424
Db 1325 TACACGAGCACCGGGTTCTTAGCTGAGTACTCTCTCTACGACTCCCAACGAGCCGCTGCC 1384

QY 1425 GGGATGTTCAATGTGCAAGACTGGAACGGTGTGATCGGAAGGAACTGCGCTGCGAGCGGTGG 1484
Db 1385 GGGCAGTTCACTGCGGCGGCGGCTGATCGGAAGGAGCTGCGCTGATGGGTGG 1444

QY 1485 GCAGCTGCGCGGATTTATGATGATGAGCGTTACTTGCAGTGAATGCCATGCCACCCAGCTTC 1544
Db 1445 GCGCATGTCACCGACACAGCGATGAGCTCAATGCAAGTTGCGAGCGCGGCGCCAGCTTC 1504

QY 1545 ACCTGCAAAAACCAAGTTCTGCAAGCCCTCTTCTGGGTCTGTGACAGTGTCAACGACTGT 1604
Db 1505 ACCTGCAAAAACCAAGTTCTGCAAGCCCTCTTCTGGGTCTGTGACAGTGTCAACGACTGT 1564

QY 1605 GGGAGCGGAGTTCAGGAGGAGGCTGCGAGCTGTCTCTGCTGGGAGTTTCAAGTGTTCGAAT 1664
Db 1565 GGGAGCAACAGCAGCAGGAGGAGGCTGCGAGTTGCTCCGCGCAGACCTTTCAGGTGTTCGAAT 1624

QY 1665 GGGAGGTGTCTCTCTCAGAGCCAGAGTGTAAATGGGAAGCAACAACTGTGGAGATGGGTCT 1724
Db 1625 GGGAGGTGTCTCTCAGAGCCAGAGTGTAAATGGGAAGCAACAACTGTGGAGATGGGTCT 1684

QY 1725 GAGAGGCTTCTATGTGACAGCGTGAATGTGCTCTTGTGACCAAAATATATCTTACCGGTGC 1784
Db 1685 GAGAGGCTTCTTCTGCGCCCAAGGTGAACGCTGCTCACTTTGTATACCAAAACACACTTACCGGTGC 1744

QY 1785 CAAATGGGCTCTGTCTGAGCAGAGGAGGAGCCCTGAGTGTGATGGGAGAGCGGAGCTGTAGC 1844
Db 1745 CTAATGGGCTCTGTCTGAGCAGAGGAGGAGCCCTGAGTGTGACGGGAGGAGGAGCTGTAGC 1804

QY 1845 GATGGCTTCCCATGAGAAAAAATGTGACTGTGGGCTGCGATCTCTTATCAAAACAGGCTCGC 1904
Db 1805 GAGGCTCAGATGAGAAAGGAGTGGAGTGTGGGCTGCGGCTCATTTACGAGACAGGCTCGT 1864

QY 1905 GTGGTTGGTGGCAGCAATGCGAGCGGCGAGTGGCCCTTGGCAGGTTGAGCTCCAGGCC 1964


```

Db 1865 GTTGTGGGGCCAGGATGGGATAGGGGAGTGGCCCTGGCAGGTAAGCCTGCATGCT 1924
QY 1965 CTGGCCAGAGGCGCACTGTGTGGGCGCTCGCTCATCTCTGCTGACTGGCTGCTCTGCA 2024
Db 1925 CTGGCCAGAGGCGCACTGTGTGGGCGCTCGCTCATCTCTGCTGACTGGCTGCTCTGCA 1984
QY 2025 GCTCATGTCTTTCCAGGATGACAAAATTTCAAGTACTTCAGACTACACGATGTGGACGGCC 2084
Db 1985 GCACACTGCTACATCGATGACAGAGGATTCAGGTACTTCAGACCCACGAGTGGACGGCC 2044
QY 2085 TTCTGGGCTGTGGACAGAGCAAGCGAGTGCCTCTGTGGGTGACAGGCTGAAGCTC 2144
Db 2045 TTCTGGGCTGTGGACAGAGCAAGCGAGTGCCTCTGTGGGTGACAGGCTGAAGCTC 2104
QY 2145 AAACGATATCATCCACCCCTCTCTCAATGATTTCACTTTCAGCTTTCAGCTATGACATCGCCTG 2204
Db 2105 AAGCGATATCATCCACCCCTCTCTCAATGATTTCACTTTCAGCTATGACATCGCCTG 2164
QY 2205 CTGAGCTGGAGAGTCCGTTGAGTACAGCACCGTCTGTGGCCCGCATCTGCTGCTGTAT 2264
Db 2165 CTGAGCTGGAGAGTCCGTTGAGTACAGCACCGTCTGTGGCCCGCATCTGCTGCTGTAT 2224
QY 2265 GCTACCATCTCTTCCCTGTGGCAAGGCCATCTGGGTACAGGCTGGGGGACACAAA 2324
Db 2225 GCTTCCCATCTCTTCCCTGTGGCAAGGCCATCTGGGTACAGGCTGGGGGACACCCCAG 2284
QY 2325 GAGGAGCTACCGGAGCGCTGATCTCAGAGGGTGAGATCGTGTCATCAACCGAGCC 2384
Db 2285 TATGGAGGCTAGTGGCGCTGATCTCTGAAAGGGTGAGATCGCGTCTCAACCGAGCC 2344
QY 2385 ACTGTGAGGACCTCAATGCGGAGCAGATCACCCACGAATGATGTGGGTTCCTC 2444
Db 2345 ACTGTGAGGACCTCAATGCGGAGCAGATCACCCCGGCGATGATGCGTGGGCTTCCTC 2404
QY 2445 AGTGGGGTGTGACTCTCCAGGGTGACTCTGGTGGCCCTTGTCAAGCGGAGAAA 2504
Db 2405 AGCGGGCGGTGACTCTCCAGGGTGATTCGGGGGACCCCTGTCCAGCGTGGAGGG 2464
QY 2505 GATGGCGGAATGTTCCAGGCTGTGTGTGCTGAGCTGGGGTGAAGGCTCGCTCAGAGAAC 2564
Db 2465 GATGGCGGAATGTTCCAGGCTGTGTGTGCTGAGCTGGGGTGAAGGCTCGCTCAGAGAAC 2524
QY 2565 AAGCAGCGGTGTACAAAGGCTCCCTGTAGTGTGGACTGTGATCAAGAGCACTGGG 2624
Db 2525 AAGCAGCGGTGTACAAAGGCTCCCTGTGTGTGGACTGTGATCAAGAGCACTGGG 2584
QY 2625 GTATAGCAGCATGACAGACAGCGGACCAACACCCACAGGATGCCCGACATGCACA 2684
Db 2585 GTATAGGCGCGGGG---GCCACCCAAATGTGTACCTCTGGGGGCCACCCATCGTCCACC 2641
QY 2685 CCTGGATACAGGAGAGAACACTGACGACATTTATGCTGTGGGCTCCCGCCCGCCACACA 2744
Db 2642 CCAGTGTGCAGC---CTGCGAGCTGGAGACTGGACCGCTGACTGCACCGAGCGCC---CCAGA 2699
QY 2745 ACCGAGCTGTGACTGCTCTTAGGACTCAGAGT 2780
Db 2700 ACATACACTGTGAATCAATCTCCAGGGCTCCAAAT 2735

```

RESULT 7

AAH23601

ID AAH23601 standard; DNA; 3147 BP.

XX AC

XX AC

XX AC

XX 03-AUG-2001 (first entry)

XX XX

XX Human TADG-15 coding sequence.

XX DE

XX XX

XX Human; TADG-15; cytostatic; vaccine; ovarian tumour; cancer;

XX KW

XX tumour antigen-derived gene 15; extracellular serine protease; ds.

XX XX

```

OS Homo sapiens.
XX Key
XX Location/Qualifiers
XX CDS
XX /*tag= a
XX /product= "Human TADG-15"
XX WO200129056-A1.
XX PN
XX 26-APR-2001.
XX
XX 20-OCT-2000; 2000WO-US29095.
XX PF
XX 20-OCT-1999; 99US-0421213.
XX PR
XX (UYAR-) UNIV ARKANSAS.
XX PA
XX O'Brien TJ, Tanimoto H;
XX PI
XX WPI; 2001-381031/40.
XX DR
XX P-PSDB; AAB98500.
XX
XX Novel extracellular serine protease, termed tumor antigen-derived gene
XX 15 protein overexpressed in carcinomas and DNA encoding it, for
XX diagnosis, treatment, prevention of cancer, particularly breast,
XX ovarian cancer
XX
XX Claim 2; Fig 2; 130pp; English.
XX
XX The present sequence is the coding sequence for human tumour
XX antigen-derived gene 15 (TADG-15) protein. TADG-15 is an extracellular
XX serine protease. It was found that TADG-15 is over-expressed in ovarian
XX tumours. TADG-15 protein or its fragments of 9-20 residues that lack
XX TADG-15 protease activity are useful for vaccinating an individual
XX against TADG-15, having, suspected of having or at risk of getting
XX cancer. Furthermore, the TADG-15 gene can be used as a diagnostic or
XX therapeutic target in cancer.
XX
XX Sequence 3147 BP; 654 A; 952 C; 958 G; 583 T; 0 other;

```

```

Query Match 60.6%; Score 1883.2; DB 22; Length 3147;
Best Local Similarity 81.2%; Pred. No. 0;
Matches 2223; Conservative 0; Mismatches 508; Indels 5; Gaps 3;
QY 45 GATCGGACCGCCCAAAACCATGGGTAGCAATCGGGGCCGCAAGCCCGAGGGGCTCTCAG 104
Db 5 GAGCGCTCGTGGGTACCATGGGAGCGATCGGCCCGCAAGGGCGAGGGGCCGGAAG 64
QY 105 GACTTCGGCGCGGAGACTCAAGTACAACTCCGGCTAGAGAAATGAATGGCTTTGAGGAG 164
Db 65 GACTTCGGCGCGGAGACTCAAGTACAACTCCGGCACGAGAAAGTGAATGGCTTTGAGGAA 124
QY 165 GGTGTGAGTCTCTGCTCGGACATGCAAGAAAGTGGAGAGCGAGGCCCGCGC 224
Db 125 GCGTGGAGTCTCTGCTCGGACATGCAAGAAAGTGGAGAGTGGCCCGGGGCGC 184
QY 225 TGGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 284
Db 185 TGGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 244
QY 285 CTGGTGTGGCACTTCCATTATCGGAATGTGGGGTTCAAAAAGTCTTCAATGGCCATCTG 344
Db 245 CTGGTGTGGCACTTTCAGTACCGGACGTGCGGTGTCCAGAAGGCTTTCATGGCTACATG 304
QY 345 AGGATCACAAATGAGATCTTTCTGGATGGTATGAGAACTCCACCTCCACAGAGTTTATC 404
Db 305 AGGATCACAAATGAGAAATTTGTGGATGCTCTCAGAGAACTCCAACTCCACTGAGTTTGA 364
QY 405 AGCTTGGCCAGCGAGGTGAAGGAGGCGCTGAAGCTGCTGTACAAATGAAGTCCCTGTCTTG 464
Db 365 AGCTTGGCCAGCAAGGTGAAGGAGGCGCTGAAGCTGCTGTACAGCGAGTCCCATTCCTG 424
QY 465 GGTCCCTACCAACAGAAAGTCCGCTGTAACTGCTTCAGTGAGGGCAGTGTCTATCGCCTAC 524

```

Db 425 GGGCCCTTACCAAGAGTGTGGCTGTGAGCGCTTACGAGGGCAGCGTCAATCGCTTAC 484
QY 525 TACTGGTTCAGAGTTCAGCATCCGCCCAACATCGGCGAGAGAGTTGATCGCGCCATGGCT 584
Db 485 TACTGGTTCAGAGTTCAGCATCCGCCCAACATCGGCGAGAGAGTTGATCGCGCCATGGCT 544
QY 585 GTGGAGCGAGTGTAAATGCAATGCAATGCAATGCAATGCAATGCAATGCAATGCAATGCA 644
Db 545 GAGGAGCGAGTGTAAATGCAATGCAATGCAATGCAATGCAATGCAATGCAATGCAATGCA 604
QY 645 TCTGTGGTGGCTTCCCATTTGACCCCGAGATGCTGAGAGGACTCAGAGCAAAAGCTGTC 704
Db 605 TCAGTGGTGGCTTCCCATTTGACCCCGAGATGCTGAGAGGACTCAGAGCAAAAGCTGTC 664
QY 705 AGTTTGTGCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 764
Db 665 AGCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 724
QY 765 AACAGTCCCTACCGGGGCTATGCGCGTTCAGTGGGTCTGCGGGGGGACGCGGACTCT 824
Db 725 GACAGCCCTTACCGGGGCTATGCGCGTTCAGTGGGTCTGCGGGGGGACGCGGACTCT 784
QY 825 GTGCTGAGCTTACCTTCCGAGCTTTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 884
Db 785 GTGCTGAGCTTACCTTCCGAGCTTTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 844
QY 885 CTGGTCACTGCTGATGATGAGCTTACCGCCCTGAGAACCCACGCTGCTGCTGCTGCTGCT 944
Db 845 CTGGTCACTGCTGATGATGAGCTTACCGCCCTGAGAACCCACGCTGCTGCTGCTGCTGCT 904
QY 945 GGCACCTTCTCACCTTCTCAAACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1004
Db 905 GGCACCTTCTCACCTTCTCAAACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 964
QY 1005 ACGCTGATAACCAATACGAGCGGAGCATCGTGGCTTTGAGGCGCACTTCTTTCAGCTG 1064
Db 965 ACCTGATTAACCAATACGAGCGGAGCATCGTGGCTTTGAGGCGCACTTCTTTCAGCTG 1024
QY 1065 CCCAAGATGAGCAGCTGCGGGCTTTTGTAGTGACACCCCAAGGAGCAATTTAGCAGCCCC 1124
Db 1025 CCTAGGATGAGCAGCTGCGGGCTTTTGTAGTGACACCCCAAGGAGCAATTTAGCAGCCCC 1084
QY 1125 TACTATCCAGCCACTACCGCCCAACATCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1184
Db 1085 TACTATCCAGCCACTACCGCCCAACATCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1144
QY 1185 AACCGGAACGTAAGGTGCGCTTCAAACTTCTTATCTGCTGCTGCTGCTGCTGCTGCTGCT 1244
Db 1145 AACCGGAACGTAAGGTGCGCTTCAAACTTCTTATCTGCTGCTGCTGCTGCTGCTGCTGCT 1204
QY 1245 GGCTCTCTGCAACCAAGGACTAGTGGAGATCAACGGGGAGAAAGTACTGCGTGAGAGGTCC 1304
Db 1205 GGCACTTCCCAAGGACTAGTGGAGATCAACGGGGAGAAAGTACTGCGTGAGAGGTCC 1264
QY 1305 CAGTTTGTGTGAGCAGCAACAGAGCAAGATTAAGTCACTTCCATCTGATCACTGCTGCT 1364
Db 1265 CAGTTTGTGTGAGCAGCAACAGAGCAAGATTAAGTCACTTCCATCTGATCACTGCTGCT 1324
QY 1365 TACAGGACACCGGGTTCCTAGCTGAGTACCTTCTTACCTGCTGCTGCTGCTGCTGCTGCTGCT 1424
Db 1325 TACAGGACACCGGGTTCCTAGCTGAGTACCTTCTTACCTGCTGCTGCTGCTGCTGCTGCTGCT 1384
QY 1425 GGGATGTTTCATGTGCAAGACTGGAACGGTGTGATCCGAAAGGAACTGCGTGTGCGAGCGCTGG 1484
Db 1385 GGGCAGTTTCATGTGCAAGACTGGAACGGTGTGATCCGAAAGGAACTGCGTGTGCGAGCGCTGG 1444
QY 1485 GCAGCTGCGCGGATTAAGTGTGAGGCTTACTGCGGATGCAATGCCAACCACAGTTTC 1544
Db 1445 GCCAGCTGCAACCGGATTAAGTGTGAGGCTTACTGCGGATGCAATGCCAACCACAGTTTC 1504
QY 1545 ACCTGCAAAACCAAGTTCGCAAGCCCTTCTTGGGCTCTGTGACAGTGTCAACGACTGT 1604

Db 1505 ACGTCGAAGAACAAAGTTCTGCAAGCCCTCTTCTTGGGTCTGCGACAGTGTGAAACGACTGC 1564
QY 1605 GGGACACGAGTGTACGAGGAGGCTGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1664
Db 1565 GGGACACGAGTGTACGAGGAGGCTGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1624
QY 1665 GGGAAAGTGTCTTCCCTCAGAGCCAGAGGTGTAATGGGAGGACAACTGTGGAGATGGGTCT 1724
Db 1625 GGGAAAGTGTCTTCCCTCAGAGCCAGAGGTGTAATGGGAGGACAACTGTGGAGATGGGTCT 1684
QY 1725 GACAGAGCTTCATGTGACAGCGTGAATGTGCTCTTGTGACACCAATATATCTACCGTGC 1784
Db 1685 GACAGAGCTTCATGTGACAGCGTGAATGTGCTCTTGTGACACCAATATATCTACCGTGC 1744
QY 1785 CAAATGGCTCTCTCTGAGCAAGGGCAACCTGAGTGTGATGGGAGAGACGAGCTGTAGC 1844
Db 1745 CTCAATGGCTCTCTCTGAGCAAGGGCAACCTGAGTGTGATGGGAGAGAGGACTGTAGC 1804
QY 1845 GATGGCTCCGATGAGAAATACTGTGACTGTGGGTGCGATCTTTTACCAACAGGCTGCG 1904
Db 1805 GACGGCTCAGATGAGAAAGGACTGCGACTGTGGGTGCGATCTTTTACCAACAGGCTGCG 1864
QY 1905 GTGGTGGTGGCAGCAATGCGGACGAGGGGAGTGGCCCTGGCAGGTGAGGCTCCACGCC 1964
Db 1865 GTTGTGGGGGACGAGTGGCGATGAGGGGAGTGGCCCTGGCAGGTGAGGCTCCACGCC 1924
QY 1965 CTGGGCGAGGGGCACTTGTGTGGGGCTCTGCTCATCTCTCTGACTGCTGCTGCTGCTGCTGCA 2024
Db 1925 CTGGGCGAGGGGCACTTGTGGGTGCTCTGCTCATCTCTCTGACTGCTGCTGCTGCTGCTGCTG 1984
QY 2025 GCTCATTTGCTTTTCCAGGATGACAAATACTTCAAGTACTCAGACTACAGTGTGGACGGCC 2084
Db 1985 GCACACTGTACATCGATGACAGAGGATTCAGGTACTCAGACCCCAACGAGTGGACGGCC 2044
QY 2085 TTCTGCGGTCTGTGGGACGAGCAAGCGCAGTGCCTCTGGGGTGGCAGGAGCTCAAGCTC 2144
Db 2045 TTCTGCGGTCTGTGGGACGAGCAAGCGCAGTGCCTCTGGGGTGGCAGGAGCTCAAGCTC 2104
QY 2145 AACGATATCATCAACCCCTTCTTCAATGATTTTCACTTCTGACTATGACTGCGCTTG 2204
Db 2105 AAGGCTATCATCTCCACCCCTTCTTCAATGATTTTCACTTCTGACTATGACTGCGCTG 2164
QY 2205 CTGAGCTGAGAGTGTGGTGGAGTACAGCAACCGTCTGCTGCGCCCATCTGCTGCTGCTGAT 2264
Db 2165 CTGAGCTGAGAGTGTGGTGGAGTACAGCAACCGTCTGCTGCGCCCATCTGCTGCTGCTGAT 2224
QY 2265 GCTACCCATCTTCTTCTGCTGCGCAAGCCATCTGGGTTCAGAGCTGCGGGGACACAAAA 2324
Db 2225 GCCTCCCATGTCTTCTTCTGCGGCAAGCCATCTGGGTTCAGAGCTGCGGGGACACCCAG 2284
QY 2325 GAGGAGGTACCGGAGCGCTGATCTTGCAGAAAGGTGAGATCCGCTGCTCATCAACGAGCC 2384
Db 2285 TATGGAGCTTGGCGCTGATCTTGCAGAAAGGTGAGATCCGCTGCTCATCAACGAGCC 2344
QY 2385 ACCTGTAGAGCTTATGCGCGAGAGATCAACCCACAGATGATGTGTGGTTCCTCCT 2444
Db 2345 ACCTGTAGAGCTTATGCGCGAGAGATCAACCCACAGATGATGTGTGGTTCCTCCTCCT 2404
QY 2445 AGTGGGGTGTGACTCTGCGAGGGTGACTCTGTGTGGCCCTTGTCAAGCGGGGAGAAA 2504
Db 2405 AGCGCGCGGTGAGACTCTTCTGCGGGGTGATTTCTGCGGGGACCCCTGCTCCAGGTGAGGCG 2464
QY 2505 GATGGGCGAATGTTCCAGGCTGTGTGTGTGAGCTGGGGTGAAGGCTGCGCTCAGAGGAAC 2564
Db 2465 GATGGGCGAATGTTCCAGGCTGTGTGTGTGAGCTGGGGTGAAGGCTGCGCTCAGAGGAAC 2524
QY 2565 AAGCGAGCTGTATACAGAGCTCCCTGTAGTTGCGGACTGTGATCAAGAGCAGACTGGG 2624
Db 2525 AAGCGAGCTGTATACAGAGCTCCCTGTAGTTGCGGACTGTGATCAAGAGCAGACTGGG 2584
QY 2625 GTATGAGCAGTGTGACAGAGCGGACCAACCAACCCACAGGAGTGCCTGATGACACA 2684
Db 2585 GTATGAGGCGCGG---GCCACCAAAATGTGTACACTGCGGGGCGCCACCATCTGTCACC 2641

QY	2685	CCTGGATACAGGAGAGGAAACACACATTTATGCTGTGGCTCCCCCCCCCAACACA	2744			
Db	2642	CCAGTGTGCAG-CCTCAGGCTGGAGACTGGACCGCTGACTGCACACGCGCC-CCAGA	2699			
QY	2745	ACCAGACTGTGAAGTGCATCTCTTAGGACTCAGAGT	2780			
Db	2700	ACATACACTGTGAAGTGCATCTCTTAGGACTCAGAGT	2735			
RESULT 8						
AAH23609/c						
ID	AAH23609 standard; RNA; 3147 BP.					
XX	AAH23609;					
XX	03-AUG-2001 (first entry)					
DT	Human TADG-15 antisense RNA sequence.					
DE	TADG-15; cytostatic; vaccine; ovarian tumour; cancer; human; antisense;					
KW	tumour antigen-derived gene 15; serine protease; ss.					
KW	Homo sapiens.					
OS	WO200129056-A1.					
XX	26-APR-2001.					
PN	20-OCT-2000; 2000WO-US29095.					
XX	20-OCT-1999; 99US-0421213.					
XX	(UYAR-) UNIV ARKANSAS.					
PA	O'Brien TJ, Tanimoto H;					
XX	WPI; 2001-381031/40.					
XX	Novel extracellular serine protease, termed tumor antigen-derived gene					
PT	15 protein overexpressed in carcinomas and DNA encoding it, for					
PT	diagnosis, treatment, prevention of cancer, particularly breast,					
PT	ovarian cancer					
XX	Example 8; Page 102-103; 130pp; English.					
PS	The present invention relates to human tumour antigen-derived gene 15					
XX	(TADG-15) protein and coding sequence (see AAH23601 and AAB98500).					
CC	TADG-15 is an extracellular serine protease. It was found that TADG-15 is					
CC	over-expressed in ovarian tumours. TADG-15 protein or its fragments of					
CC	9-20 residues that lack TADG-15 protease activity are useful for					
CC	vaccinating an individual against TADG-15, having, suspected of having or					
CC	at risk of getting cancer. Furthermore, the TADG-15 gene can be used as a					
CC	diagnostic or therapeutic target in cancer. TADG-15 gene can be used as					
CC	expressed in the opposite orientation such that an antisense RNA molecule					
CC	(the present sequence) was produced. The present sequence is useful for					
CC	hybridising to the complementary RNA in a cell and thereby inhibiting					
CC	translation of TADG-15 into protein.					
XX	Sequence 3147 BP; 583 A; 958 C; 952 G; 654 U; 0 other;					
SQ	Query Match 60.6%; Score 1883.2; DB 22; Length 3147;					
Best Local Similarity 81.2%; Pred. No. 0;						
Matches 2223; Conservative 0; Mismatches 508; Indels 5; Gaps 3;						
QY	45	GATCGGACGCCAAACCATGGTACATCGGGCCGCAAGCCGGAGGGGCTCTGAG	104			
Db	3143	GACGGCCTCGGGTACCATGGGAGCGATCGGGCCCGCAAGCGGGGGCCCGAAG	3084			
QY	105	GACTTCGGCGCGGACTCAAGTAACTCCGGCTAGAGAAATGAATCGCTTTGAGGAG	164			
Db	3083	GACTTCGGCGGGACTCAAGTAACTCCGGCAGAGAAAGTGAATGCTTTGAGGAA	3024			

Db	863	TATGGAGGCACTGGCGCGCTGATCTCTGC AAAAGGTTGAGATCGCGGTCAACAACGAGCC	804
QY	2385	ACCTGTGAGAACCTCATGCGCAGCAGATCACCCACGAAATGATGTGTGGGTTTCCTC	2444
Db	803	ACCTGCGAGAACCTCTCGCGCAGCAGATCACCCGCGCATGATGTGCGTGGGCTTCCTC	744
QY	2445	AGTGGGGGTGTGACTCCTCCACGGGTGACTCTGTGTGGGCCCTTGTCAAGCGCGGAGAAA	2504
Db	743	AGCGGGCGGTGGACTCCTCGCAGGGTGATTCGGGGGACCCCTGTCCACGCTGGAGCG	684
QY	2505	GATGGGGGAATGTTCCAGGCTGTGTGTGGTGAAGGCTGCGCTCAGAGGAAC	2564
Db	683	GATGGGCGGATCTTCCAGGCGGTGTGTGGTGAAGGCTGCGCTCAGAGGAAC	624
QY	2565	AAGCAGGGGTGTGACTCCTCCACGGGTGACTCTGTGTGGGCCCTTGTCAAGCGCGGAGAAA	2624
Db	623	AAGCCAGGCGGTGTGACTCCTCCACGGGTGACTCTGTGTGGGCCCTTGTCAAGCGCGGAGAAA	564
QY	2625	GTATAGCAGCATGGACAGCAGCGCGACCAACACACCCACAGGGGATGCCGACATGCACA	2684
Db	563	GTATAGGGGCGGG---GCCACCAATGTGTACACTGGGGGCCACCATCGTCCACC	507
QY	2685	CCTGGATACAGGAGAGGAACATGACAGCATTTATGCTGTGGCTCCCCCCCCCAACACA	2744
Db	506	CCAAGTGTGCAGC-CCTGCGGGCTGGAGACTGGACCGCTGACTGCACGAGGCC-CCAGA	449
QY	2745	ACCAGACTGTGAATGCATCCTTAGGACTCAGAT	2780
Db	448	ACATACACTGTGAATCAATCTCCAGGGCTCCAAAT	413
RESULT 9			
AAL53444			
ID	AAL53444 standard; DNA; 3147 BP.		
XX	AAL53444;		
XX	12-DEC-2002 (first entry)		
XX	Type II transmembrane serine protease 1 coding DNA SEQ ID No 1.		
DE	Cytostatic; type-II membrane-type serine protease 7; MTSP7; malignancy; neoplastic disease; pre-malignant lesion; tumour; zymogen form; cancer; malignant; enzyme; gene; ds.		
XX	Homo sapiens.		
OS	WO200272786-A2.		
PN	19-SEP-2002.		
PD	13-MAR-2002; 2002WO-US07903.		
XX	13-MAR-2001; 2001US-275592P.		
XX	(CORV-) CORVAS INT INC.		
XX	Madison EL, Ong EO;		
PI	WPI; 2002-732827/79.		
XX	P-PSDB; AA022929.		
DR	New transmembrane serine protease 7 (MTSP7) polypeptide for diagnosing neoplastic diseases, monitoring tumor progress or therapeutic effectiveness, or identifying MTSP7 modulators for treating tumors or cancers		
PT	Disclosure; Page 168-172; 184pp; English.		
XX	The invention relates to a purified single or two-chain polypeptide, which comprises the protease domain of a type-II membrane-type serine protease 7 (MTSP7) or its catalytically active portion. The polypeptide comprising MTSP7 is useful for detecting or diagnosing a neoplastic disease.		

QY	1965	CTGGCCAGGGCCACTGTGTGGGGCTCGCTCACTCTCTGACTGGCTGGTCTCTGCA	2024
Db	1925	CTGGCCAGGGCCACATCTCGGTGCTTCCCTCATCTCTCCAACTGGCTGTCTGTCC	1984
QY	2025	GCTCATTTGCTTTTTCAGGATGACAAAATTTTCAAGTACTCAGACTACAGATGTGGACGGCC	2084
Db	1985	GCACACTGTCTACATCGATGACAGAGGATTTAGGTACTCAGACCCACGCGATGGACGGCC	2044
QY	2085	TTCTCTGGGTCTGTGGACCAAGAGCAAGCGCAGTGCCTCTGCGGTGTCAGGAGCTGAAGCTC	2144
Db	2045	TTCTCTGGGTCTGACGACACAGAGCCAGCGCAGCGCCCTTGGGTGTCAGGAGCGCAGCTC	2104
QY	2145	AAAGGTATCATCACCCACCCTTCTCTTCAATGATTTTCACTTCGACTATGACATCGCCCTTG	2204
Db	2105	AAGCGCATCATCTCCCAACCCCTTCTTCAATGACTTCACTTCGACTATGACATCGCGCTG	2164
QY	2205	CTGGAGCTGGAGAGTCCGTGGAGTACAGCACCGTCTGTGCGCCCACTCTGCTGCGCTGAT	2264
Db	2165	CTGGAGCTGGAGAAAACGGGCAGAGTACAGTCTCATGTGCGGCCATCTGCTTCCGCGAC	2224
QY	2265	GCTAACCATGTCTTCCCTGTGTGGCAAGGCCATCTGGGTCAAGGCTGGGGGCACACAAA	2324
Db	2225	GCCTCCCATGTCTTCCCTGCGGCAGAGCCCATCTGGGTCAAGGCTGGGGGCACACCCAG	2284
QY	2325	GAGGAGGTACCGGAGCGCTGATCTTGCAAGAGGTGAGATCCGTGTCTCATCAACGACACC	2384
Db	2285	TATGGAGGCACTGGCGGCTGATCTTGCAAAAGGTGAGATCCGCGTCAACAACGACACC	2344
QY	2385	ACTGTGAGGACCTCATGCGCGCAGCAGATCACCCCAAGATGATGTGTGGGTTCCTC	2444
Db	2345	ACTGCGAGAACTCTCTGCGCAGCAGATCACGCCGCGCATGATGTGCGTGGCTTCCTC	2404
QY	2445	AGTGGGGGTGTGACTCTCTGCGCAGGTGACTCTGTGTGGCCCTTTGTCAAGCGCGGAGAAA	2504
Db	2405	AGCGCGGGGTGTGACTCTCTGCGCAGGTGATCTCGGGGGGACCCCTGTCCAGCGTGGAGGG	2464
QY	2505	GATGGGCGAATGTTCCAGGTGTGTGTGTGAGCTTGGGGTGAAGSGCTCGCTCAGAGGAC	2564
Db	2465	GATGGCGGATCTTTCAGCGCGTGTGGTGTGAGCTGGGAGACGGCTCGCTCAGAGGAAAC	2524
QY	2565	AAACCGAGGCTGTACACAAGGCTCCCTGTAGTTTCGGGACTGGATCAAGAGCACACTGGG	2624
Db	2525	AAAGCCAGGGGTGTACACAAGGCTCCCTCTGTTTCGGGACTGGATCAAGAGAACACTGGG	2584
QY	2625	GTATAGCAGCATGGAACAGACAGCGGACCAACAACCCACAGGGATGCCGACATGCACA	2684
Db	2585	GTATAGGGCCGGG-----GCCACCAAAATGTGTACACTCGGGGGCCACCATCGTCCACC	2641
QY	2685	CCTGGATACAGGAGAGGAACACTGACGACATTTATGCTGTGGCTCCCCCCCCCAACACA	2744
Db	2642	CCAGTGTGCACG-CTGTGAGGCTGGAGACTGGACCGCTGACTGCACACAGGGCCC-CCAGA	2699
QY	2745	ACCCGACTGTGNACTGTGATCTCTTTAGGACTCAGAGT	2780
Db	2700	ACATACACTGTGNACTCAATCTCCAGGGGTCCAAAT	2735

RESULT 10

AAAL53445

ID AAL53445 standard; DNA; 3147 BP.

AC AAL53445;

XX
DT 12-DEC-2002 (first entry)

XX
DE Type II transmembrane serine protease 1 domain DNA SEQ ID No 3.
....

XX Cytostatic; type-II membrane-type serine protease 7; MTP7; malignancy;;
 KW neoplastic disease; pre-malignant lesion; tumour; zymogen form; cancer;
 KW malignant; enzyme; gene; ds.

XX
SO
XX

PN WO200272786-A2.
 XX
 PD 19-SEP-2002.
 XX
 PF 13-MAR-2002; 2002WO-US07903.
 XX
 PR 13-MAR-2001; 2001US-275592P.
 XX
 PA (CORV-) CORVAS INT INC.
 XX
 PI Madison EL, Ong EO;
 XX
 PI WPI; 2002-732827/79.
 DR P-PSDB; AAO22930.
 DR
 XX
 PT New transmembrane serine protease 7 (MSP7) polypeptide for diagnosing
 PT neoplastic diseases, monitoring tumor progress or therapeutic
 PT effectiveness, or identifying MSP7 modulators for treating tumors or
 PT cancers -
 XX
 PS Disclosure; Page 174-176; 184pp; English.
 XX
 CC The invention relates to a purified single or two-chain polypeptide,
 CC which comprises the protease domain of a type-II membrane-type serine
 CC protease 7 (MSP7) or its catalytically active portion. The polypeptide
 CC comprising MSP7 is useful for detecting or diagnosing a neoplastic
 CC disease, a pre-malignant lesion, a malignancy or other pathologic
 CC condition in a subject. This polypeptide is also useful for monitoring
 CC tumour (e.g. tumor of the breast, cervix, prostate, lung, ovary or
 CC colon) progress and/or therapeutic effectiveness. The inhibitor of the
 CC polypeptide or the inhibitor of the polypeptide's zymogen form is useful
 CC for treating or preventing a neoplastic disease, or tumour initiation,
 CC growth or progression, or a (pre-)malignant condition. The polypeptide or
 CC polynucleotide is also useful for identifying modulators of MSP7, which
 CC may be used to treat cancers or tumours. This polynucleotide sequence
 CC represents the coding DNA for a protein domain of the type-II membrane-
 CC type serine protease 1 relating to the invention.
 XX
 QQ Sequence 3147 BP; 654 A; 952 C; 958 G; 583 T; 0 other;

Db 425 GGGCCCCACCAAGGAGTGGGCTGTGACGGCCCTTACGGGGGACGGTCAATCGCCTAC 484
QY 525 TACTGGTCAGAGTTACAGATCCCCCCACACCTGGCAGAGAGTTGATCGGGCCATGGCT 584
Db 485 TACTGGTCTGAGTTACAGATCCCCCAGACCTGGTGAGAGGGCCGAGCGCTCATGGCC 544
QY 585 GTGAGGAGGTTGTAACATTGCGCACCCCGAGCAGCGGCATGAAATCTTGGTCTTAACA 644
Db 545 GAGGAGGCGGTAGTATGCTGCCCCCGCGGGCGGCTCCCTGAAAGTCTCTTTGGTCAAC 604
QY 645 TCTGTGTGGCCTTCCCATTTAGCCACCCAGATGTGACAGGACTCAGGACAAAGCTGC 704
Db 605 TCAGTGTGGCTTTTCCCCACGGAATCCAAAACAGTACAGAGACCCAGGACAAAGCTGC 664
QY 705 AGTTTTCGCCCTGATGCCCATGTGTGACAGCTGACAGCTTCACTACCCCTGGCTTCCCC 764
Db 665 AGCTTTGGCTGACGCGCCCGGGTGTGAGCTGATGGCTTACCAACGCCCGGCTTCCCT 724
QY 765 AACAGTCCCTACCCGGCGCATGCCCGCTGCGAGTGGGTCTGTGCGGGGGAGCGCGACTCT 824
Db 725 GACAGCCCTTACCCCGCTCATGTCCCGCTGCGAGTGGGCGCTTGGCGGGGAGCGCGACTCA 784
QY 825 GTGCTGAGCTCACCTTCCGAAGCTTTGATGTGCTCCCTGTGATGAGCATGCGAGTGAC 884
Db 785 GTGCTGAGCTCACCTTCCGAGCTTTTGAACCTTGGCTCTGCGACGAGCGCGCAGCGAC 844
QY 885 CTGCTCACCGGTGATGATAGCTGAGCCCATGAGAACCCACAGCTGTGTGGCGGCTGTGT 944
Db 845 CTGCTGAGCGGTGTACAAACCCCTGAGCCCCATGAGCCCCACCGCTTGTGTGTGTGT 904
QY 945 GGCACCTTCTCACCTCTTACAACTGACTTCTCTCTCTCCCGAAGAGCTTCTCTGTGC 1004
Db 905 GGCACCTTACCTCCCTCCTCAACCTGACCTTCCACTCTCTCCAGAACGCTCCTGCTCATC 964
QY 1005 ACGTGATAACCAATACTGACCGCGACATCTCGGCTTTGAGGCGCATCTTCTTCCAGCTG 1064
Db 965 AACTGATAACCAACTGAGCGGCGCATCTCCGGCTTTGAGGCGCATCTTCTTCCAGCTG 1024
QY 1065 CCAAGATGAGCAGCTGTGGCGGCTTTTGGTGAAGACACCAAGGAGCATTTAGCAGCCCC 1124
Db 1025 CTTAGGATGAGCAGCTGTGGAGCGGCTTACGTAAAGCCACCGGGGACATTCACAGCCCC 1084
QY 1125 TACTATCCAGCCACTACCGCCCAACATCAATGACATGGAATATCAAGGTGCCAAC 1184
Db 1085 TACTATCCAGCCACTACCGCCCAACATTTGACTGCACATGGAATATGAGGTGCCAAC 1144
QY 1185 AACCGGAACGTGAAGGTGCGCTTCAAACTCTCTATCTGTGTGACCCCAAGTACCACTG 1244
Db 1145 AACAGCATGTGAAGGTGAGCTTCAAAATCTTCTTACCTGTGAGCCCGCGCTGCGTGG 1204
QY 1245 GGCTCTGACCAAGGACTATGTGGAGATCAACGGGGAGAACTCTCGGTGAGAGTCC 1304
Db 1205 GGCACCTGCCCCAGGACTACGTGGAGATCAATGGGAGAAATACTTGGGAGAGAGTCC 1264
QY 1305 CAGTTTGTGTGAGCAGACACGACAGCAAGATTACAGTCCACTTCCATTTGATCACTCG 1364
Db 1265 CAGTTTGTGTGAGCAGACACGACAGCAAGATTACAGTCCACTTCCACTCAGATCAGTCC 1324
QY 1365 TACACGACACCGGGTTCCTAGCTGAGTACTCTCTTACGACTCCCAAGACCCGCTGCCCA 1424
Db 1325 TACACGACACCGGGTTCCTAGCTGATACCTCTCTTACGACTCCAGTGAACCCATGCCG 1384
QY 1425 GGGATTTGATGTGCAAGACTGAGCAGGTGATCCGAAAGGAACTGCGCTGCGAGCGGTGG 1484
Db 1385 GGGCAGTTTACGTGCGCCACCGGGCGGTGTATCCGGAAGGAGCTGCGCTGTGTGAGTGG 1444
QY 1485 GCAGACTGCCCGGATATAGTGTGATGAGCTTACTGCGATGCAATGCCACCCACAGTTC 1544
Db 1445 GCCGACTGCAACCCACACAGCATGAGCTCAACTGAGTTGGCAGCGCGGCCACCACTTC 1504
QY 1545 ACGTGCAAAACCCAGTTCTGCAAGCCCTCTTCTGGGTCTGTGACAGTGTCAACGACTGT 1604

Db 1505 ACBTGCAAGAACAAAGTTCTTCAAGACCCCTCTTCTGGGTCTTGGCAGCTGTGTAACGACTGC 1564
QY 1605 GGGGACGGAAGTGCAGGAGGGCTGCAGCTGTCTCTGCTGGGAGTTTCAAGTGTTCGAAT 1664
Db 1565 GGAACAACAGCAGCAGGAGGGGTGCAGTTGTTCGGCCCGAGACCTTTCAGGTGTTCGAAT 1624
QY 1665 GGGAAAGTGTCTCCCTCTCAGAGCCAGAAAGTGTAAATGGGAAGGACAACTGTGGAGATGGGTCT 1724
Db 1625 GGGAAAGTGCCTCTTCGAAAACCCAGCAGTGCATTTGGGAGAGCAGCTGTGGGGACGGGTCC 1684
QY 1725 GACAGGCTTCATGTGACAGCGTGAATGTCTCTTTGCAACCAATAATACTACCGCTGC 1784
Db 1685 GACGAGGCTCTCTGCCCCCAGAGTGAACGTCTCCTTGTACCAACACACACTTACCGCTGC 1744
QY 1785 CAAAATGGGCTCTGTCTGACCAAGGCGAACCTCAGTGTGATGGGAGAGCGGACTGTAGC 1844
Db 1745 CTCAATGGGCTCTGTGTGACCAAGGCGAACCTCAGTGTGACGGGAGGAGGACTGTAGC 1804
QY 1845 GATGGCTCCGATGAGAAAACTGTGACTGTGGCTCGATCTCTTTTACCAACAGGCTCGC 1904
Db 1805 GACGGCTCAGATGAGAAGGACTTCGACTGTGGCTGGGTCTTTCAGGACAGAGCTCGT 1864
QY 1905 GTGGTTGGTGGCAGCAATGCGGACGAGGGGAGTGGCCCTGGCAGAGTGAGCCCTCGAGCC 1964
Db 1865 GTTGTGGGGGCGACGATGCGGATGAGGGCGAGTGGCCCTGGCAGGTAAAGCTGTGCT 1924
QY 1965 CTGGGCCAGGGCCACTTGTGTGGGGCTCTGCTCATCTCTCTGACTGGCTGTCTGTGA 2024
Db 1925 CTGGGCCAGGGCCACTTGTGGGGCTTCCCTCATCTCTCCCACTGGCTGTGCTCTGCC 1984
QY 2025 GCTCATTTGCTTTCCAGGATGACAAAAATTTCAAGTACTCAGACTTACAGATGTGGAGCGCC 2084
Db 1985 GCACACTGCTACATCGATGACAGAGGATTACAGTACTCAGACCCACGCGATGCGAGCC 2044
QY 2085 TTCTCTGGTCTGTGACACGACAGCAGCGAGTGCCTCTGGGGTGCAGGAGCTGAAGCTC 2144
Db 2045 TTCTCTGGGCTTGCACACGACGAGCCAGCGCCCTGGGGTGCAGGAGCGCAGGCTC 2104
QY 2145 AAAGTATCATACCCACCTCTCTTCAATGATTTCACTTGCAGCTTACAGATGTGGAGCGCTG 2204
Db 2105 AAGGGATCATCTCCACCCCTTCTTCAATGACTTCACTTGCAGTATGACATCGCGCTG 2164
QY 2205 CTGGAGCTGGAGAGTTCGGTGGAGTACAGCACCGTCTGTGGCGCCCATCTGCTCTGCCTGAT 2264
Db 2165 CTGGAGCTGGAGAAACCGGACAGTACAGCTCCATGTGTGGCGCCCATCTGCTCTGCCGAC 2224
QY 2265 GCTACCCATGTCTTCTCTGCTGGCAGGCGCATCTGGGTACAGGCTGGGGGACACAAA 2324
Db 2225 GCTCTCCATGTCTTCTCTGCTGGGCAAGGCGCATCTGGGTACAGGCTGGGGACACCCAG 2284
QY 2325 GAGGAGGTACCGGAGCGCTGATCTCTGCAAGAGGTGAGATCCGTCTCATCAACAGACC 2384
Db 2285 TATGGAGGCACTGGCGCTGATCTCTGCAAGAGGTGAGATCGCGCTCATCAACAGACC 2344
QY 2385 ACCTGTGAGGACCTCATGCGGACGAGATCACCCCAAGAAATGATGTGTGGTTCCTC 2444
Db 2345 ACCTGTGAGAACTCTTCTGCGCAGCAGATCACGCGCGCATGATGTGCTGGCTTCTCTC 2404
QY 2445 AGTGGGGTGTGGACTCTCTGCGAGGCTGACTGTGGTGGCCCTTGTCAAGCGGGAGAA 2504
Db 2405 AGCGGGCGGTGGACTCTTGGCAGGGGTGATCTCGGGGGACCCCTGTCTCCAGCGTGGAGCG 2464
QY 2505 GATGGCGCAATGTCTCAGGCTGTGTGGTGGCTGGGTGAAAGGCTGCGCTCAGAGGAAC 2564
Db 2465 GATGGCGGATCTTCCAGGCGGTGTGTGGTGGCTGGGGAGCGGTGCTGCTCAGAGGAAC 2524
QY 2565 AAGCAGGCGGTGTACAAAGGCTCTCTGTGTGGGACTGGATCAAGAGACACTGGG 2624
Db 2525 AAGCAGGCGGTGTACAAAGGCTCTCTGTGTGGGACTGGATCAAGAGAACTGGG 2584
QY 2625 GTATAGCAGTGTGACAGCAGCGCACCAACACCCACAGGGATGCCGACATGCACA 2684
Db 2585 GTATAGGCGCGGG---GCCACCCAAATGTGTACACCTGCGGGGCGACCCATCGTCCACC 2641

QY 2685 CCTGGATCAGGAGGAGAACATGACGACATTTATCTGTGGCTCCCGCCCAACACA 2744
 DB 2642 CCAGTGTGACG-CCTGCGAGGTGGAGACTGGACCGCTGACTGCACGAGGCC-CCAGA 2699
 QY 2745 ACCGACTGTGAAGTGCATCTTCTAGGACTCAGAGT 2780
 DB 2700 ACATACACTGTGAACATCTCCAGGGCTCCAAAT 2735

RESULT 11

ABZ58500

ID ABZ58500 standard; cDNA; 3147 BP.

XX AC ABZ58500;

XX DT 13-MAY-2003 (first entry)

XX DE Transmembrane serine protease 1 (MTSP1) cDNA.

XX KW Human; transmembrane serine protease 1; MTSP1; MTSP20; enzyme;
 XX KW Cystostatic; dermatological; cardiant; vulnerary; ophthalmological;
 XX KW gene therapy; gene; ss.

XX OS Homo sapiens.

XX FH Key Location/Qualifiers
 XX FT CDS 23..2590

XX FT /*tag= a

XX FT /product= "MTSP1"

XX PN W02003004681-A2.

XX PD 16-JAN-2003.

XX PF 03-JUL-2002; 2002WC-US21208.

XX PR 03-JUL-2001; 2001US-302939P.

XX PA (CORV-) CORVAS INT INC.

XX PI Madison EL, Ong EO;

XX DR WPI; 2003-239207/23.

XX PS P-PSDB; ABP72376.

XX PT New type-II membrane-type serine protease 20 polypeptides, useful for
 XX PT preparing a medicament for diagnosing, treating or preventing cancer,
 XX PT dermatological disorders, aberrant wound repairs or crest syndromes -
 XX PS Disclosure; Page 194-198; 216pp; English.

CC The present sequence is that of cDNA encoding human type II
 CC transmembrane serine protease 1 (MTSP1). The invention relates
 CC to novel human type II transmembrane serine protein 20 (MTSP20)
 CC (see ABP72374) and nucleic acids encoding it (see ABZ58499).
 CC Also claimed are methods of inhibiting tumour initiation, growth
 CC or progression by inhibiting MTSP20 activity, and of treating or
 CC preventing a disease or disorder associated with undesired and/or
 CC uncontrolled angiogenesis or neovascularisation, especially
 CC undesired angiogenesis associated with solid neoplasms, vascular
 CC malformations and cardiovascular disorders, chronic inflammatory
 CC diseases, aberrant wound repairs, circulatory disorders, crest
 CC syndromes, dermatological disorders and ocular disorders using
 CC an inhibitor of MTSP20.

XX SQ Sequence 3147 BP; 654 A; 952 C; 958 G; 583 T; 0 other;

Query Match 60.6%; Score 1883.2; DB 25; Length 3147;
 Best Local Similarity 81.2%; Pred. No. 0;
 Matches 2223; Conservative 0; Mismatches 508; Indels 5; Gaps 3;

QY 45 GATCGGACCGCCAAACCATGGGTAGCACTGGGCGCCGCAAGGCCGAGGGGCTCTCAG 104

DB 5 GAGCGCCCTCGGGGTACCATGGGAGCGATCGGGCCCGCAAGGGCGAGGGGCCCCGAAG 64
 QY 105 GACTTCGGCGGGGACTCAAGTACAACTCCCGGCTAGAGAAATGAATGGCTTTGAGGAG 164
 DB 65 GACTTCGGCGGGGACTCAAGTACAACTCCCGGCGAGAGAAATGAATGGCTTTGAGGAG 124
 QY 165 GGTGTGGAGTTCTCGCTCGGAACAAATGCCAAGAAAGTGGAGAGCGAGGCGCCAGCGC 224
 DB 125 GCGGTGGAGTTCTCGCCAGTCAACAACTGCAAGAGGTGGAAAGCATGGCCCGGGGCGC 184
 QY 225 TGGGTGGTGTCTGGTGGAGTGTCTTTCAGCTTCCTTGTCTCTCCCTCAATGGCTGGCTTG 284
 DB 185 TGGGTGGTGTCTGGCAGCGGTGTCTGATCGGCTCTCTTGGTCTTGTGGGATCGGCTTC 244
 QY 285 CTGGTGTGSCACTTCCATTTATCGGAATGTGGGGTTCAAAAGTCTTCAATGGCCATCTG 344
 DB 245 CTGGTGTGCAATTTGCAGTACCGGACGCTGCTGTCAGAGGTCTTCAATGGCTACATG 304
 QY 345 AGGATCACAAATGAGATCTTTCTGATCGTATGAGAACTCCACCTCCACAGATTTATC 404
 DB 305 AGGATCACAAATGAGATTTTGTGATGCTTACGAGAACTCCAACTCCACTGAGTTGTA 364
 QY 405 AGCTTGGCCAGCCAGGTGAAGAGGCGCTGAAAGTGTCTGTACAATGAAGTCCCTGTCTG 464
 DB 365 AGCTTGGCCAGCAAGGTGAAGGACGCGCTGAAAGTGTCTGTACAGCGGAGTCCCATTC 424
 QY 465 GGTCTCTACCAAGAAAGTTCGCTGTAAGTGTCTTCAAGTGGGAGGCTCATCCCTAC 524
 DB 425 GGGCCCTACCAAGAGGTGCGCTGTGAGGCGCTTCAAGGAGGCGGCTCATCCCTAC 484
 QY 525 TACTGTGAGATTTCAGATCCCCCACCTGGCAGAGAGGTTGATCGGCCCATGGCT 584
 DB 485 TACTGTGCTGATTCAGCATCCCGCAGCACCTGTGTGGAGGAGCGCGCTCATGGCC 544
 QY 585 GTGGAGCGAGTTGTAAATTTGGCCACCCCGCAGCACCGGCACTGAAATCTTCTGTCTA 644
 DB 545 GAGGAGCGCGTAGTCACTGTCCCGCGGGCGGCTCTCTGAAAGTCTCTTGTGTCA 604
 QY 645 TCTGTGTGGCTTCCCATTTAGACCCCAAGATGCTGCGAGAGACTCAGAGCAACAGCTGC 704
 DB 605 TCAGTGTGGTGTTCCTCCCAAGCACTCCAAACAGTACAGAGAGCCCAAGCAACAGCTGC 664
 QY 705 AGTTTGGCTTGCATGCCCCATGGTGCAGAGTGAACACGCTTCACTACCCCTGGCTTCCC 764
 DB 665 AGCTTTGGCTTGCACGCGCGCGGTGTGGAGTGTATGCTTCAACACGCGCTTCCCT 724
 QY 765 AACAGTCCCTACCGGCGCATGCGGCTGCCAGTGGTCTTGGGGGAGAGCCGCACTCT 824
 DB 725 GACAGCCCTTACCGCGCTCATGCGCGCTGCGAGTGGGCGCTTGGGGGGGAGCGCACTCA 784
 QY 825 GTGTGTGAGCTTCACTTCCGAAAGCTTTGATGTGCTCTCTGTGATGAGCATGGCAGTGC 884
 DB 785 GTGTGTGAGCTTCACTTCCGAGCTTTGACCTTGGTCTTGGAGAGCGCGGAGCGAC 844
 QY 885 CTGTGTACCGTGTATGATGAGTCCGAGCCCAAGAAACCCACGCTGTGTGGTGGCTGTGT 944
 DB 845 CTGTGTACCGTGTATCAACACCTGAGCCCACTGAGCCCAAGCGCTGTGTGTGTGT 904
 QY 945 GGCACCTTCAACCTTCACTCAACCTGCTTCTCTCTCTCCAGAGCTTCTTCTGTCTC 1004
 DB 905 GGCACCTTCACTTCTCTCTCAACCTGACCTTCCACTCTCTCCAGAGCTCTGTCTCATC 964
 QY 1005 ACGTGTATAACCAATATCTGACCGGCGGACATCTTGGCTTTGAGGCGCACTTTTTCAG 1064
 DB 965 ACCTGTATAACCAACACTGAGCGGGGCGCATCCCGGCTTTGAGGCGCACTTTTCCAG 1024
 QY 1065 CCCAGATGAGCAGTGTGGCGGCTTTTGTAGTGACACCCCAAGGAGCATTTAGCAGCCCC 1124
 DB 1025 CCTAGGATGAGCAGTGTGGAGGCGGCTTACGTAAAGCCCAAGGCGGACATTCACAGCCCC 1084
 QY 1125 TACTATCCAGGCGCACTACCGCCCAACATCACTGCACTGGAATATCAAGGTGCCCAAC 1184

1085 TACTACCCAGCCACTACCCACCCAAACATTGACATGCAATGGAACATTGAGGTGCCAAC 1144
1185 AACCGGAACGTGAAGTGCCTTTCAAACTCTTCTATCTGTGTGACCCCAACACGTACCAAGTG 1244
1145 AACACGATCTGAAGTGTGAGCTTCAAAATCTTCTACCTGTGAGCCCGGCTGCTGG 1204
1245 GGCTCTGACACCGAGTATGTGAGATCAACGGGAGAGTACTGCGGTGAGAGTCC 1304
1205 GGCACCTGCCCCAAGGACTACGTGGAGATCAATGGGAGAAATACTGCGGAGAGAGTCC 1264
1305 CAGTTTGTGTGAGCAGCAACAGCAGCAAGATTACAGTTCACACTTCCTCATCTGATCACTCG 1364
1265 CAGTGTGTGTCTACACGACCAACAGCAACAGATCACAGTTGCGTTCACACTGATCACTGTC 1324
1365 TACACGACACCGGGTCTTCTAGTGTAGTACTCTCTCTACAGTCTCAACACGACCGGTGCCA 1424
1325 TACACCGACACCGGCTCTTCTAGTGTAAATCTCTCTCTACAGTCTCACTGAGTCCAGTCCCG 1384
1425 GGGATGTTTCACTGTGCAAGACTGCAAGCTGCATCGGAAGAACTGCGCTCGAGAGGCTGG 1484
1385 GGGAGTTTACGTGCGCGCACGGGGCGGTGTATCCGGAAGAGAGTGGCTGTGATGGCTGG 1444
1485 GCAGACTGCCCGGATTATAGTGTAGTGTAGTGTACTGTGCGGATGCAATGCAACACACAGTTC 1544
1445 GCGCACTGCAACCGACACAGCATGAGCTCACTGAGTTGCAAGCGCGGCAACAGTTC 1504
1545 ACCTGCAAAACACAGTTCTGCAAGCCCTCTTCTGAGTGTGTGACAGTGTCAACAGTGT 1604
1505 ACCTGCAAAACACAGTTCTGCAAGCCCTCTTCTGAGTGTGTGACAGTGTCAACAGTGT 1564
1605 GGGACCGGAAGTCAACGAGAGGCTGCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1664
1565 GGNAGCAACAGCGACAGCAGAGGGGTGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1624
1665 GGAAGTGTCTCCTCAGACCCAGAGTGTGTGGAAGGACAACTGTGTGAGATGAGTCT 1724
1625 GGAAGTGTCTCCTCAGAAAGCCAGCAGTGTGATGGAGGACGACACTGTGGGACGGGTCC 1684
1725 GACGAGCTTCATGTGACAGCTGGAATGTGCTCTTGTGACCAAAATATACCTACCGCTGC 1784
1685 GACGAGCTTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1744
1785 CAAATGGCTCTGCTGAGCAGGCGCAACCTGAGTGTGATGGAGGACGACGACTGTGAG 1844
1745 CTCATGGCTCTGCTGAGCAGGCGCAACCTGAGTGTGATGGAGGACGACGACTGTGAG 1804
1845 GATGCTCCGATGAGAAAATCTGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1904
1805 GACGCTCAGATGAGAAGGACTGCGACTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1864
1905 GTGGTGTGTGACAGTGTGACAGGAGGCGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1964
1865 GTTGTGGGGGACCGATCGGATGAGGCGGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1924
1965 CTGGGCGAGGGCCACTGTGTGGGGCTCGCTCATCTCTCTGCTGCTGCTGCTGCTGCTGCT 2024
1925 CTGGGCGAGGGCCACTGTGGGGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1984
2025 GTCATGTGCTTTCAGGATGACAAAATTTCAAGTACTCAGACTACAGCTGTGAGCGGCC 2084
1985 GCACACTGTACATCGATGACAGAGGATTCAAGTACTCAGACTACAGCTGTGAGCGGCC 2044
2085 TTCTTGGGTCTGTGACAGCAGCAGCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2144
2045 TTCTTGGGTCTGTGACAGCAGCAGCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2104
2145 AAACGTATCATCAACCCACCTCTCTTCAATGATTTCACTTCACTTCACTTCACTTCACTT 2204
2105 AAGCGATCATCTCCACCCCTTCTTCAATGATTTCACTTCACTTCACTTCACTTCACTT 2164
2205 CTGGAGCTGGAGAGTCTGCTGAGTACAGCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2264
2165 CTGGAGCTGGAGAAACCGCAGAGTACAGCTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2224

2265 GCTACCCATGTCTTCCCTGTGSCAAGGCACTCTGGGTACAGGCTGGGGGACACAAAA 2324
2225 GCTCCCATGTCTTCCCTGTGSCAAGGCACTCTGGGTACAGGCTGGGGGACACCCAG 2284
2325 GAGGAGGTACCGGACGCTGTATCTCTGCAAGAGGTGAGATCCGTGTCTCATCAACAGACC 2384
2285 TATGAGGCACTGGCGCGTGTATCTCTGCAAGAGGTGAGATCCGTGTCTCATCAACAGACC 2344
2385 ACCTGTGAGGACCTCATGCGCGCAGCAGATCAACCCACGATGATGTGTGTGGTTCCTC 2444
2345 ACCTGTGAGGAACTCTCTGCGCAGCAGATCAACCCACGATGATGTGTGTGGTTCCTC 2404
2445 AGTGGGGGTGTGAGTCTCTGCGCAGGCTGACTCTGTGTGGCCCTTGTCAAGCGGGAGAAA 2504
2405 AGCGCGCGGTGGACTCTCTGCGCAGGCTGATTCGGGGGACCCCTGTCTCAGCGTGGAGCG 2464
2505 GATGGCGCAATGTTCCAGGCTGTGTGAGTGTGGGTGAAAGGTGCGCTCAGAGGAAC 2564
2465 GATGGCGGATCTCTCAGCGCGGTGTGTGAGTGTGGGGAGACGGCTGCGCTCAGAGGAAC 2524
2565 AAGCCAGGCGGTGTACACAAAGGCTCCCTGTGTGAGTGTGGGACTGGATCAAGAGAACACTGG 2624
2525 AAGCCAGGCGGTGTACAAAGGCTCCCTGTGTGAGTGTGGGACTGGATCAAGAGAACACTGG 2584
2625 GTATAGCAGATGGACAGACAGCCGACCAACCAACACAGGATGCCCGACATGCACA 2684
2585 GTATAGGCGCGGG--GCCACCAAAATGTGTACACTGCGGGGCCACCCATCGTCCACC 2641
2685 CTTGATACAGAGAGAGAACACTGACGACATTTATGTGTGGCTTCCCCCCCCCAACACA 2744
2642 CCAAGTGTGACG--CCTGACAGGCTGGAGACTGACCGCTGACTGCACAGCGCCC-CCAGA 2699
2745 ACCGAGCTGTCAACTGCACTCTTAGGACTCAGAGT 2780
2700 ACATACCTGTGAACTCAATCTCCAGGGCTCCAAAT 2735

RESULT 12

ABZ58501

ID ABZ58501 standard; cDNA; 3147 BP.

AC ABZ58501;

XX XX

DT 13-MAY-2003 (first entry)

XX XX

Transmembrane serine protease 1 (MTSP1) protein domain cDNA.

Human; transmembrane serine protease 1; MTSP1; MTSP20; enzyme;

cytostatic; dermatological; cardiant; vulnery; ophthalmological;

gene therapy; gene; ss.

XX OS

Homo sapiens.

XX Key

Location/Qualifiers

1865..2590

/*tag= a

/product= "MTSP1 protease domain"

XX WO2003004681-A2.

XX PD

16-JAN-2003.

XX XX

03-JUL-2002; 2002WO-US21208.

XX XX

03-JUL-2001; 2001US-302939P.

XX XX

(CORV-) CORVAS INT INC.

XX PI

Madison EL, Ong EO;

XX XX

WPI; 2003-239207/23.

XX P-PSDB; ABF72377.

XX New type-II membrane-type serine protease 20 polypeptides, useful for
PT preparing a medicament for diagnosing, treating or preventing cancer,
PT dermatological disorders, aberrant wound repairs or crest syndromes -
XX
PS Disclosure; Page 200-202; 216pp; English.
XX
CC The present sequence is that of cDNA encoding the protease domain
CC of human type II transmembrane serine protease 1 (MTSP1). The
CC invention relates to novel human type II transmembrane serine protein
CC 20 (MTSP20) (see ABP72374) and nucleic acids encoding it (see
CC AB258499). Also claimed are methods of inhibiting tumour initiation,
CC growth or progression by inhibiting MTSP20 activity, and of treating
CC or preventing a disease or disorder associated with undesirable and/or
CC uncontrolled angiogenesis or neovascularisation, especially
CC undesired angiogenesis associated with solid neoplasms, vascular
CC malformations and cardiovascular disorders, chronic inflammatory
CC diseases, aberrant wound repairs, circulatory disorders, crest
CC syndromes, dermatological disorders and ocular disorders using an
CC inhibitor of MTSP20.
XX
SQ Sequence 3147 BP; 654 A; 952 C; 958 G; 583 T; 0 other;

Query Match 60.6%; Score 1883.2; DB 25; Length 3147;
Best Local Similarity 81.2%; Pred. No. 0;
Matches 2223; Conservative 0; Mismatches 508; Indels 5; Gaps 3;

QY 45 GATCGACCGCCAAACCATGGTAGCAATCGGGCCGCAAGCGCGAGGGGGCTCTCAG 104
DB 5 GAGCGGCTCGGGGTACCATGGGAGCGATCGGGCCCGCAGGCGGAGGGCCCGNAG 64

QY 105 GACTTCGGCGGGGACTCAAGTACACTCCCGGTAGAGAACATGAATGGCTTTAGAGAG 164
DB 65 GACTTCGGCGGGGACTCAAGTACACTCCCGGTAGAGAACATGAATGGCTTTAGAGAA 124

QY 165 GGTGTGGAGTTCTCTGCTCGTCAAAATGCCAAGAGTGGAGAGCGGCCCGCAGGGC 224
DB 125 GGCCTGGAGTTCTCTGCTCGTCAAAATGCCAAGAGTGGAGAGCGGCCCGCAGGGC 184

QY 225 TGGGTGGTGGTGGGAGTCTGTTCAGCTTCCTCTTCCTCTCTCTCTCTCTCTCTCT 284
DB 185 TGGGTGGTGGTGGGAGTCTGTTCAGCTTCCTCTCTCTCTCTCTCTCTCTCTCTCT 244

QY 285 CTGCTGGGACCTTCATATCGGAATGTGGGGTTCAAAAGTCTTCAATGGCCATCTG 344
DB 245 CTGCTGGGACCTTCATATCGGAATGTGGGGTTCAAAAGTCTTCAATGGCTACATG 304

QY 345 AGGATCAAAATGAGATCTTTCTCGATGCGTATGAGAACTCCACTCCACAGAGTTTATC 404
DB 305 AGGATCAAAATGAGATTTTGTGGATGCGCTGAAGCTGTGTACAGAGTCCCACTCC 364

QY 405 AGCTGCGCCAGCGAGGTGAAGGAGCGCTGAAGTGTCTGTACATGAAGTCCCTGTCTG 464
DB 365 AGCTGCGCCAGCGAGGTGAAGGAGCGCTGAAGTGTCTGTACAGAGTCCCACTCC 424

QY 465 GGTCTCTACACAGAGTGGCTGTAACTGCTTCTAGTAGGGCAGTGTATCGCCTAC 524
DB 425 GGTCTCTACACAGAGTGGCTGTAACTGCTTCTAGTAGGGCAGTGTATCGCCTAC 484

QY 525 TACTGTGTACAGTTTACATCTCCCGCCACACTGCGCAGAGAGTTGATCGGCCATGGCT 584
DB 485 TACTGTGTACAGTTTACATCTCCCGCCACACTGCGCAGAGAGTTGATCGGCCATGG 544

QY 585 GTGAGCGAGTTGTAAATTCGACACCGGACAGCGGCACTGAAATCTTCTGTCTAAACA 644
DB 545 GTGAGCGAGTTGTAAATTCGACACCGGACAGCGGCACTGAAATCTTCTGTCTAAACA 604

QY 645 TCTGTGTGGCTTCCCGTTCAGAGTGTGAGAGTGTGAGAGTGTGAGAGTGTGAGAGT 704
DB 605 TCTGTGTGGCTTCCCGTTCAGAGTGTGAGAGTGTGAGAGTGTGAGAGTGTGAGAGT 664

QY 705 AGTTTTCCTGATGCCATGTGTGACAGTGTGACAGCTTCACTACCTCCCTGGCTTCCC 764
DB 705 AGTTTTCCTGATGCCATGTGTGACAGTGTGACAGCTTCACTACCTCCCTGGCTTCCC 764

DB 665 AGCTTTGGCTGCACGCGCCGCGGTGTGAGAGTGTGAGCTTCAACACGCGCCGCTTCCCT 724
QY 765 AACAGTCCCTACCGCGGCGATGCGCGCTGCCAGTGGTCTTGGGGGGGACGCCGACTCT 824
DB 725 GACAGCCCTTACCGCGGCTCATGCGCGCTGCCAGTGGGCGCTTGGGGGGGACGCCGACTCA 784
QY 825 GTGCTGAGCCTCACCTTCCGAAGCTTTGATGTGCTGCTCTGTGATGAGCATGGCAGTGAC 884
DB 785 GTGCTGAGCCTCACCTTCCGAAGCTTTGATGTGCTGCTCTGTGATGAGCATGGCAGTGAC 844
QY 885 CTGCTACCGTGTATGATGAGCCTTCCGAAGCTTTGATGTGCTGCTCTGTGATGAGCATGG 944
DB 845 CTGCTGAGCCTCACCTTCCGAAGCTTTGATGTGCTGCTCTGTGATGAGCATGGCAGTGAC 904
QY 945 GGCACCTTCTCACCTCTCAACCTGACCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1004
DB 905 GGCACCTTCTCACCTCTCAACCTGACCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 964
QY 1005 ACGTGTATTAACCAATACTGACCGGCGCATCTCTGCTTTGAGGCGCATCTTCTTCCAGCTG 1064
DB 965 ACGTGTATTAACCAATACTGACCGGCGCATCTCTGCTTTGAGGCGCATCTTCTTCCAGCTG 1024
QY 1065 CCCAAGATGAGCAGTGTGGCGGCTTTTGTAGTGACACCCCAAGGAGCAATTTAGCAGCCCC 1124
DB 1025 CCTAGGATGAGCAGTGTGGAGGCGCTTTACGTAAAGCCAGGCGCATTTCAACAGCCCC 1084
QY 1125 TACTATCCAGCCACTACCCGCCCAACATCACTGACATGCACTGCAATATCAAGTGCGCCAA 1184
DB 1085 TACTATCCAGCCACTACCCGCCCAACATTTGACTGCACTGCAATATCAAGTGCGCCAA 1144
QY 1185 AACCGAAGCTGAAGTGGCTTCAAACTCTTCTTATCTGTGTGACCCCAAGCTACAGTG 1244
DB 1145 AACCGAAGCTGAAGTGGCTTCAAACTCTTCTTATCTGTGTGACCCCAAGCTACAGTG 1204
QY 1245 GGCTCTTCCACCAAGGACTATGTGGAGTCAACCGGGAGAGTATCTGGGGTGAAGGTTC 1304
DB 1205 GGCACCTGCCCCAAGGACTAGTGGAGTCAATGGGGAGAAATACTGCGGAGAGAGTTC 1264
QY 1305 CAGTTTCTGTGTGAGCAGCAACAGCAGCAAGATACAGTCCACTTCCATCTGTACTCTG 1364
DB 1265 CAGTTTCTGTGTGAGCAGCAACAGCAGCAAGATACAGTCCACTTCCATCTGTACTCTG 1324
QY 1365 TACACGACACCGGGTTCTTAGTGTAGTACCTCTCTTACGACTCCAAAGACCCGCTGCCCA 1424
DB 1325 TACACGACACCGGGTTCTTAGTGTAGTACCTCTCTTACGACTCCAGTACCCATGCGCG 1384
QY 1425 GGGATGTTCATGTGCAAGACTGACGCTGATCGAAAGAACTGCGCTGCGAGCGGTGG 1484
DB 1385 GGGCAGTTTCACTGCGCGCAGCGGGCGGTGTATCCGGAAGGAGCTGCGCTGTGATGGTGG 1444
QY 1485 GCAGACTGCGCGGATTTATAGTGTAGTACCTTCTTACGACTCCAAAGACCCGCTGCCCA 1544
DB 1445 GCGCAGTGCACCGACACAGCGATGAGTCACTGCACTGCGAGCGCGCGCCACAGTTC 1504
QY 1545 ACGTCAAAACACAGTTCTGAAGCCCTCTTCTGGGTCTGTGACAGTGTCAACGACTGT 1604
DB 1505 ACGTCAAAACACAGTTCTGCAAGCCCTCTTCTGGGTCTGTGACAGTGTGAACGACTGC 1564
QY 1605 GGGGACGGAAGTGCACGAGGAGGCTGCGCTGCTCTGGGAGTTTCAAGTGTTCAT 1664
DB 1565 GGGGACGGAAGTGCACGAGGAGGCTGCGCTGCTCTGGGAGTTTCAAGTGTTCAT 1624
QY 1665 GGGGAGTGTCTCCCTCAGAGCCAGAGAGTGTAAATGGGAAGGACAACTGTGGAGATGGTCT 1724
DB 1625 GGGGAGTGTCTCTGAAAGCCAGAGTGCATGGAAGAGACGACTGTGGGGACGGGTTC 1684
QY 1725 GACGAGCTTCATGTGACAGCGTGAATGTGCTCTTGTGACCAAAATATACCTACCGCTGC 1784
DB 1685 GACGAGCTTCCTGCCCCAAGTGAACGCTGCTTGTGTGACCAAAACACACCTTACCGCTGC 1744
QY 1785 CAAATATGGCTCTGTGAGCAAGGGCAACCTGTAGTGTGATGGGAAGACGCACTGTAGC 1844
DB 1745 CTAATATGGCTCTGTGAGCAAGGGCAACCTGTAGTGTGATGGGAAGGAGGACTGTAGC 1804

QY	1845	GATGGCTCCGATGAGAAAACTGTGACTGTGGCTGGATCCTTTATCCAAACAGGCTCGC	1904
Db	1805	GACGGCTCAGATGAGAAGGACTGGACTGTGGCTGGCGATTTTCAGAGACAGGCTCGT	1864
QY	1905	GTGGTTGTGGCAGGAATGCGGACGAGGGCGAGTGGCCCTGGCAGGTGAGCTTCCAGCC	1964
Db	1865	GTGTGTGGGGCAGCGATGCGGATGAGGGCGAGTGGCCCTGGCAGGTAAAGCTGCATGCT	1924
QY	1965	CTGGGCCAGGGCCACTTGTGTGGGCGCTCGCTCATCTCTCTGACTGGCTGGTCTCTGCA	2024
Db	1925	CTGGGCCAGGGCCACATCTGCGGTGCTTTCCCTCATCTCTCCAACTGGCTGGTCTCTGGC	1984
QY	2025	GCTCATTTCTTTTCAGGATGACAAAAATTTCAAGTACTCAGACTACACGATGTGGACGGCC	2084
Db	1985	GCACACTGCTACATCGATGACAGAGGATTCAAGTACTCAGACCCTCAGCTAGTGGACGGCC	2044
QY	2085	TTCTTGGGTCTGTGGACGAGCAAGCGCAGTGCCTCTGGGTGTCAGGAGCTGAAGCTC	2144
Db	2045	TTCTTGGGTCTGTGGACGACGAGCCAGCGAGCGCCCTGGGGTGCAGGAGCGCAGCTC	2104
QY	2145	AAACGTATCATCACCCACCCCTTCCTTCAATGATTTACCTTCGACTATGATCATCGCGCTG	2204
Db	2105	AAGCGCATCATCTCCACCCCTTCTTCAATGACTTCACTTCGACTATGATCATCGCGCTG	2164
QY	2205	CTGAGCTTGGAGAAAGTCCGTGGAGTACAGCACCGCTCGTGGCCGCCATCTCCCTGCCCTGAT	2264
Db	2165	CTGAGCTTGGAGAAACCGGCAGAGTA CAGCTTCATGTGCGGCCCATCTCCCTGCCGAC	2224
QY	2265	GCTACCCATGCTTTCCTCTGTGGCAAGGCCATCTTGGGTCA CAGCTTGGGGGACACAAAA	2324
Db	2225	GCCTCCCATGCTTTCCTCTGGCGGCAAGGCCATCTTGGGTCA CAGCGCTGGGGACACACCAG	2284
QY	2325	GAGGAGGTACCGGAGCGCTGATCCTCGAAGAGGTGAGATCCGTGTCTATCAACACGAGCC	2384
Db	2285	TATGAGGCACTGGCGGCTGATCTCTGCAAGAGGTGAGATCCGCTGTCATCAACACGAGCC	2344
QY	2385	ACTGTGAGAACCTCATGCCGACGAGATCACCCCAAGATGATGTGTGGGTTCCTTC	2444
Db	2345	ACCTGCAGAACTCTCTGCCGACGAGATCACGCCGCGCATGATGTGCGTGGGCTTCTTC	2404
QY	2445	AGTGGGGTGTGACTCTCTGCAAGGTGACTCTGTGGGCCCTTTGTCAAGCGCGGAGAAA	2504
Db	2405	AGCGCGCGCTGACTCTCTCCACAGGTGATTTCCGGGGGACCCCTGTCCAGCGTGGAGGCG	2464
QY	2505	GATGGGCGAATGTTCCAGGCTGTGTGGTCAGCTGGGGTGAAGCGTCGCTCAGAGGAC	2564
Db	2465	GATGGGCGGATCTTCCAGGCGGTGTGGTGAGCTGGGAGACGGCTCGCTCAGAGGAC	2524
QY	2565	AAGCAGGCGTGTACACAAGGCTCCCTGTAGTTTCCGACTGGATCAAGAGACACACTGGG	2624
Db	2525	AAGCAGGCGTGTACACAAGGCTCCCTCTGTTTCGGGACTGGATCAAGAGAACACTGGG	2584
QY	2625	GTATAGCAGCATGGACAGACAGCCGACCAACAACACACAGGGATGCCCGACATGCACA	2684
Db	2585	GTATAGGGGCGGG---GGACCCAAATGTGTACCTCGGGGGCCACCATGCTCCACC	2641
QY	2685	CCTGGATACAGGAGAGGAAACATCAGACATTTATGCTGTGGCTCCCCCCCCCACAACA	2744
Db	2642	CCAGTGTGCAGC--CTTGAGGCTTGGAGACTGGACCGCTGACTGCACACAGCGCCC--CCAGA	2699
QY	2745	ACCAGACTGTGAATGTCATCTCTTAGACTTCAGAGT	2780
Db	2700	ACATACACTGTGAATCAATCTCCAGGGTCCAAAT	2735

RESULT 13
ABZ22450
ID ABZ22450 standard; cDNA; 3147 bp.
XX
XX ABZ22450;
XX
DT 24-MAR-2003 (first entry)

XX	Human membrane-type serine protease MTSP1 encoding cDNA SEQ ID NO:1.
DE	Human; membrane-type serine protease; enzyme; MTSP10; cytostatic;
XX	type-II membrane-type serine protease; neoplastic disease; tumour;
KW	MTSP1; matriptase; gene; ss.
KW	
OS	Homo sapiens.
XX	
XX	Key Location/Qualifiers
EH	23..2590
FT	CDS
FT	/*tag= a
FT	/product= "membrane-type serine protease MTSP1"
FT	/note= "also known as matriptase"
XX	
XX	W0200292841-A2.
XX	
XX	21-NOV-2002.
XX	
XX	14-MAY-2002; 2002WO-US15332.
XX	
XX	14-MAY-2001; 2001US-291001P.
XX	
XX	(CORV-) CORVAS INT INC.
XX	
XX	Madison EL, Yeh J;
PI	
XX	WFI; 2003-129309/12.
DR	
DR	P-PSDB; ABP56619.
XX	
XX	New polypeptides comprising the protease domain of a type-II
PT	membrane-type serine protease (MTSP10), or its mucins, useful for
PT	diagnosing neoplasms or malignancies, or for screening for MTSP10
PT	inhibitors for treating such diseases
XX	
XX	Example 1; Page 177-181; 198pp; English.
PS	
XX	
CC	The present invention describes a polypeptide comprising a purified
CC	single or two chain polypeptide, which comprises the protease domain of
CC	a type-II membrane-type serine protease (MTSP10) or its catalytically
CC	active portion, or a mucin of it, where up to 50 % of the amino acids
CC	are replaced with another amino acid, and the resulting polypeptide is a
CC	single chain or two chain polypeptide that has a catalytic activity of
CC	at least 1-10 % of the unmutated polypeptide. MTSP10 has cytostatic
CC	activity. The polypeptide containing the protease domain of the MTSP10
CC	is useful for detecting a neoplastic disease, and for diagnosing the
CC	presence of a pre-malignant lesion, a malignancy, or other pathologic
CC	condition in a subject, or monitoring tumour (e.g. breast, cervix,
CC	prostate, lung, ovary or colon tumour) progression and/or therapeutic
CC	effectiveness. An inhibitor of the polypeptide containing the protease
CC	domain of MTSP10 is useful for treating or preventing neoplastic disease
CC	in a mammal. An inhibitor of the activation cleavage of the zymogen
CC	form of the MTSP10 polypeptide is useful for inhibiting tumour
CC	initiation, growth or progression, or treating (pre-)malignant conditions
CC	of the e.g. breast, cervix, prostate, lung, ovary or colon. The present
CC	sequence encodes human MTSP1 (also known as matriptase), which is used
CC	in an example from the present invention.
CC	

Sequence 3147 BP; 654 A; 952 C; 958 G; 583 T; 0 other;

Query Match 60.6%; Score 1883.2; DB 25; Length 3147;
Best Local Similarity 81.2%; Pred. No. 0;
Matches 2223; Conservative 0; Mismatches 508; Indels 5; Gaps 3

QY	45	GATCGGACCGCCAAAACCATCGGTAGCAATCGGGCCGCAAGGCCGAGAGGGGCTCTCAG	104
Db	5	GAGCGGCTTCGGGTATCCATCGGAGCGATCGGCCCCGCAAGGCGGAGGGGCCCGCAAG	64
QY	105	GACTTCGGCGGGGACTCAAGTACAACCTCCGGCTAGAGAACATGATGCTTTTGAGGAG	164
Db	65	GACTTCGGCGCGGGACTCAAGTACAATCCCGGCACGAGAAAGTGAATGCTTTGGAGGAA	124
QY	165	GGTGTGGAGTTTCCTGCCTCGCAACATGCCAAGAAAGTCGAGAGACGAGGCCCCCGAGCGC	224

QY 2385 ACCTGTGAGGACCTCATCCGCGAGCAGATCACCCACGAATGATGTGTGGTTTCTC 2444
 Dd |||||
 QY 2345 ACCTGCGAAGACCTCTCTGCGCAGCAGATCAGCGCGCATGATGTGTGGTTTCTC 2404
 Dd |||||
 QY 2445 AGTGGGGTGTGACTCTCTGCGCAGGTCATCTGTGTGGTGGCCCTTGTCAAGCGCGAGAAA 2504
 Dd |||||
 QY 2405 AGCGCGCGGTGGACTCTCTGCGCAGGTCATCTCTGCGGCGACCCCTGTCCAGCGTGGAGCG 2464
 QY 2505 GATGGCGCAATGTTTCCAGCGTGTGTGTGAGCTGGGTGAAGCTGGCTCAGAGAAC 2564
 Dd |||||
 QY 2465 GATGGCGGATCTTCCAGCGGTGTGTGTGAGCTGGGAGACGGTGGCTCAGAGAAC 2524
 QY 2565 AAGCGAGCGTGTACAAAGCTCCTCTAGTTCGGGACTGGATCAAGAGCACTGGG 2624
 Dd |||||
 QY 2525 AAGCGAGCGTGTACAAAGCTCCTCTAGTTCGGGACTGGATCAAGAGCACTGGG 2584
 QY 2625 GTATAGCAGCATGCAGACAGCGCCACCAACACACAGGATGCCGATGCCACATGCACA 2684
 Dd |||||
 QY 2585 GTATAGGCGCGGG---GCCACCCAAATGTGTACACCTGCGGGGCCACCCATCGTCCACC 2641
 QY 2685 CCTGGATACAGGAGGAGAACACTGACGACATTTATGCTGTGGCTTCCCGCCCAACACA 2744
 Dd |||||
 QY 2642 CCAGTGTGACG-CCTGAGGCTGGAGACTGGACGCTGACTGCACCGGCCCC-CCAGA 2699
 QY 2745 ACCGAGACTGTGAATCTGATCCTTTAGGACTCAGAGT 2780
 Dd |||||
 QY 2700 ACATACACTGTGAATCAATCTCCAGGCTCCAAAT 2735

RESULT 14

ABZ22451
 ID ABZ22451 standard; cDNA; 3147 BP.

XX AC ABZ22451;
 XX DT 24-MAR-2003 (first entry)
 XX DE Human MTSP1 protease domain encoding cDNA SEQ ID NO:3.
 XX KW Human; membrane-type serine protease; enzyme; MTSP10; cytosolic;
 KW type-II membrane-type serine protease; neoplastic disease; tumour;
 KW MTSP1; matriptase; gene; ss.
 XX OS Homo sapiens.

XX FH Key Location/Qualifiers
 FT CDS 1865..2590
 FT /*tag= a
 FT /partial
 FT /product= "MTSP1 protease domain"
 FT /note= "no start codon given"

XX WO200292841-A2.

XX 21-NOV-2002.

XX 14-MAY-2002; 2002WO-US15332.

XX 14-MAY-2001; 2001US-291001P.

XX (CORV-) CORVAS INT INC.

XX Madison EL, Yeh J;

XX WPI; 2003-129309/12.

XX P-PSDB; ABP56620.

XX PT New polypeptides comprising the protease domain of a type-II
 PT membrane-type serine protease (MTSP10), or its mutants, useful for
 PT diagnosing neoplasms or malignancies, or for screening for MTSP10
 PT inhibitors for treating such diseases -

XX

Example 1; Page 183-185; 198pp; English.

PS The present invention describes a polypeptide comprising a purified
 XX single or two chain polypeptide, which comprises the protease domain of
 CC a type-II membrane-type serine protease (MTSP10) or its catalytically
 CC active portion, or a mutin of it, where up to 50 % of the amino acids
 CC are replaced with another amino acid, and the resulting polypeptide is a
 CC single chain or two chain polypeptide that has a catalytic activity of
 CC at least 1-10 % of the unmutated polypeptide. MTSP10 has cytosolic
 CC activity. The polypeptide containing the protease domain of the MTSP10
 CC is useful for detecting a neoplastic disease, and for diagnosing the
 CC presence of a pre-malignant lesion, a malignancy, or other pathologic
 CC condition in a subject, or monitoring tumour (e.g. breast, cervix,
 CC prostate, lung, ovary or colon tumour) progression and/or therapeutic
 CC effectiveness. An inhibitor of the polypeptide containing the protease
 CC domain of MTSP10 is useful for treating or preventing neoplastic disease
 CC in a mammal. An inhibitor of the activation cleavage of the zymogen
 CC form of the MTSP10 polypeptide is useful for inhibiting tumour
 CC initiation, growth or progression, or treating (pre-)malignant conditions
 CC of the e.g. breast, cervix, prostate, lung, ovary or colon. The present
 CC sequence encodes the protease domain of human MTSP1 (also known as
 CC matriptase), which is used in an example from the present invention.

XX Sequence 3147 BP; 654 A; 952 C; 958 G; 583 T; 0 other;

Query Match 60.6%; Score 1883.2; DB 25; Length 3147;

Best Local Similarity 81.2%; Pred. No. 0;

Matches 2223; Conservative 0; Mismatches 508; Indels 5; Gaps 3;

QY 45 GATCGGACCGCCAAACCATGGGTAGCAATCGGGGCGCGCAAGCGCGGAGGGCTCTCAG 104
 Dd |||||
 QY 5 GAGCGGCTCTCGGGTACCATGGGAGGATCGGCGCGCAAGCGCGGAGGGCCCCGAAG 64
 Dd |||||
 QY 105 GACTTCGGCGCGGACTCAAGTACAACTCCCGCTAGAGAACATGAATGCTTTGAGGAG 164
 Dd |||||
 QY 65 GACTTCGGCGCGGACTCAAGTACAACTCCCGCTAGAGAACATGAATGCTTTGAGGAA 124
 QY 165 GGTGTGGAGTTCCTGCTCGAACAATGCCAAGAGTGGAGAGCGAGGCCCGAGCGC 224
 Dd |||||
 QY 125 GGGTGGAGTTCCTGCGCAGTCAACAACGTCAGAGGTGGAAGCATGGCGCGGCGC 184
 QY 225 TGGTGGTGTGTGTGGCAGTGTCTTTCAGTTCCTCTTCTCTCCCTCATGGCTGGCTTG 284
 Dd |||||
 QY 185 TGGTGGTGTGTGGCAGCTGTGTATCGGCTCTCTTGTCTTGTGGGATCGGCTTC 244
 QY 285 CTGGTGTGGCATTCCATTATCGAATGTGGGTTCAAAAAGTCTTCAATGGCCCATCTG 344
 Dd |||||
 QY 245 CTGGTGTGGCATTTCAGTACCGGGACGTGCTGCTCCAGAAGGTCTTCAATGGCTACATG 304
 QY 345 AGGATCACAATGAGATCTTCTTGGATGCTATGAGAACTCCACCTCCACAGAGTTTATC 404
 Dd |||||
 QY 305 AGGATCACAATGAGAAATTTTGTGGATGCTACGAGAACTCCACCTCCAGTTTGTGA 364
 QY 405 AGCTGGCCAGCCAGGTGAAGGAGCGCTGAGAGTGTCTGTACAATGAAGTCCCTGTCCTG 464
 Dd |||||
 QY 365 AGCTGGCCAGCAAGGTGAAGGAGCGCTGAGAGTGTCTGTACAAGGAGTCCCATTCCTG 424
 QY 465 GGTCCCTACCAAGAAGTCTGTAAGTGTCTGAGTGTGAGGAGGAGTGTCTATCGCTTAC 524
 Dd |||||
 QY 425 GGGCCCTACCAAGAGTCTGAGTGTGAGGAGTGTGAGGAGGAGTGTCTATCGCTTAC 484
 QY 525 TACTGTGAGAGTTCAGCATCCCGCCACACTGGCAGAGAGGTGTATCGCGCATGGCT 584
 Dd |||||
 QY 485 TACTGTGAGTTCAGCATCCCGCAGCAGTGTGTGGAGGAGGAGGAGTGTATGGCC 544
 QY 585 GTGGAGCGAGTGTAACTATGTCACCCGAGCAGCGGCACTGAAATCTCTTCTGTGTAA 644
 Dd |||||
 QY 545 GAGGAGCGGTAGTATGTGTGCGCGCGCGCTCCCTGAAGTCTCTTGTGTGTACC 604
 QY 645 TCTGTGTGGCTTCCCAATGACCCAGAGTGTGAGAGGAGTCTCAGGAGCAACAGCTGC 704
 Dd |||||
 QY 605 TCAGTGTGGCTTTCCCAACGAGTCTCCAAAACAGTACAGAGGAGCCAGGACCAACAGCTGC 664

QY 705 AGTTTGGCTGATGCCATGCTGAGAGTGAACAGCTTCACTACCCCTGGCTTCCC 764
Db 665 AGCTTGGCTGACAGCCCGGGTGGAGCTGATGCGCTTACACAGCCCGCTTCCCT 724
QY 765 AACAGTCCCTACCCCGCATCCCGCTGCCAGTGGGTCTGCGGGGAGCGCGACTCT 824
Db 725 GACAGCCCTACCCCGCTCATGCCCTGCCAGTGGGCCCTGCGGGGAGCGCGACTCA 784
QY 825 GTGCTGAGCTCACCTTCGGAAGCTTTGATGCTGCTTGTGATGAGCATGCGAGTGC 884
Db 785 GTGCTGAGCTCACCTTCGGAAGCTTTGATGCTGCTTGTGATGAGCATGCGAGTGC 844
QY 885 CTGCTCAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 944
Db 845 CTGCTCAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 904
QY 945 GGCACCTTCTCACCTTCTCAACACCTGACTTCTCTCTCTCTCTCTCTCTCTCT 1004
Db 905 GGCACCTTCT 964
QY 1005 AGCTGATTAACCAATGATGACCGCGCAATCCTGCTTTGAGGCCACTTCTTCCAGCTG 1064
Db 965 ACATGATAACCAACATGAGCGCGCATCCCGCTTTGAGGCCACTTCTTCCAGCTG 1024
QY 1065 CCCAGATGAGCGCTGCGGCTTTTGTAGTGACCCCAAGGAGATTTAGAGCGCCC 1124
Db 1025 CTTAGATGAGCGCTGCGGCTTTTGTAGTGACCCCAAGGAGATTTAGAGCGCCC 1084
QY 1125 TACTATCAGGCGCACTACCCCGCAACATCAACTGACATGCAATATCAAGGTGCCAAC 1184
Db 1085 TACTATCAGGCGCACTACCCCGCAACATCAACTGACATGCAATATCAAGGTGCCAAC 1144
QY 1185 AACCGAAGTGAAGTGGCTTCAAACTTTCTATCTGTTGAGGCCCAACGTTCCAGTG 1244
Db 1145 AACCGAAGTGAAGTGGCTTCAAACTTTCTATCTGTTGAGGCCCAACGTTCCAGTG 1204
QY 1245 GGTCTCTGACCAAGGACTATGAGAGTCAACCGGAGAGTACTGCGGTGAGAGTCC 1304
Db 1205 GGCACCTGCCCCAAGGACTATGAGAGTCAACCGGAGAGTACTGCGGTGAGAGTCC 1264
QY 1305 CAGTTTGTGTGAGCAGCAACAGCAGCAAGATTAACGTCCACTTCCATTTGATCACTCG 1364
Db 1265 CAGTTTGTGTGAGCAGCAACAGCAGCAAGATTAACGTCCACTTCCATTTGATCACTCG 1324
QY 1365 TACAGGACACCGGGTCTTAGCTGAGTACTCTCTAGCTCAACAGCCCGTGGCCA 1424
Db 1325 TACAGGACACCGGGTCTTAGCTGAGTACTCTCTAGCTCAACAGCCCGTGGCCA 1384
QY 1425 GGGATGTTCTATGTCAGAGCTGGAAGTGGTATCCGGAAGGAACTGCGTGGCAGCGTGG 1484
Db 1385 GGGCAGTTACGTCCGCAAGGGCGGTGATCCGGAAGGAGCTGCGTGTGATGGCTGG 1444
QY 1485 GCAGACTGCCCGGATTAAGTATGAGCGTTTACTGCGGATGCAATGCCACCCAGTTTC 1544
Db 1445 GCCGACTGCAACCGACACAGCATGAGCTCAACTGAGTTCGAGCGCGGCCACAGTTTC 1504
QY 1545 ACCTGCAAAACAGGTTCTGCAAGCCCTCTTCTGCGGTCTGTGAGTCAAGTCAAGCTGT 1604
Db 1505 ACCTGCAAAACAGGTTCTGCAAGCCCTCTTCTGCGGTCTGTGAGTCAAGTCAAGCTGT 1564
QY 1605 GGGCAGGAGTGAAGGAGGCGTGGAGCTGCTCTCTGCGGAGGAGCTGCGTGTGATGGCTGG 1664
Db 1565 GGGCAGGAGTGAAGGAGGCGTGGAGCTGCTCTCTGCGGAGGAGCTGCGTGTGATGGCTGG 1624
QY 1665 GGGAGTGTCTCTCTCAGAGCCAGAGTGAATGGGAAGGAGCACTGTGGAGTGGGTCT 1724
Db 1625 GGGAGTGTCTCTCTCAGAGCCAGAGTGAATGGGAAGGAGCACTGTGGAGTGGGTCT 1684
QY 1725 GACGAGGTTCTATGACAGGTTGAATGCTGCTCTCTGTCACCAAAATATACCTACCGCTGC 1784
Db 1685 GACGAGGCTCTCTGCCCCAAGGTTGAAGCTGCTGCTACTTGTACCAAAACACACCTACCGCTGC 1744
QY 1785 CAAAATGGCTCTCTCTGAGCAAGGGCAACCTCTAGTGTGATGGAGAGCGGACTGTAGC 1844

RESULT 15
AAD47180
ID AAD47180 standard; DNA; 3147 BP.
XX
AC AAD47180;

Db 1745 CTCATGGGCTCTGCTTGGCAAGGGCAACCTGAGTGTGACGGGAGGAGCTGTAGC 1804
QY 1845 GATGGCTCCGATGAGAAAACCTGTGACGTGTGGGTGCGATCTCTTTACAAACAGAGCTGCG 1904
Db 1805 GACGGCTCAGATGAGAAAGGACTGGGACTGTGGGTGCGGTCACTTACGAGAGCAGGCTCGT 1864
QY 1905 GTGGTTGTGGCAGCAATGCGGACGAGGGCGAGTGGCCCTGGCAGGTGAGCTCCACGCC 1964
Db 1865 GTTGTGGGGGCAACGGATGCGGATGAGGGCGAGTGGCCCTGGCAGGTGAGCTGCTGATGCT 1924
QY 1965 CTGGGCAAGGGCCACTTGTGTGGGGCCCTGCTCATCTCTCTGACTGTGCTGTCTCTCA 2024
Db 1925 CTGGGCAAGGGCCACTTGTGTGGGGCCCTGCTCATCTCTCTGACTGTGCTGTCTCTGCT 1984
QY 2025 GCTCATTTGCTTTCAGGATGACAAAATTTCAAGTACTCAGACTACAGCATGTGAGCGGCC 2084
Db 1985 GCACATGCTACATCATGATGACAGAGGATTTCAAGTACTCAGACTACAGCATGTGAGCGGCC 2044
QY 2085 TTCTGGGCTCTGCTGGACACGAGCAAGCGAGTGCCTCTGGGGTGCAGAGCTGAGCTC 2144
Db 2045 TTCTGGGCTCTGCTGGACACGAGCAAGCGAGTGCCTCTGGGGTGCAGAGCTGAGCTC 2104
QY 2145 AAACGATATCATCAACCCCTTCTTCAATGATTTCACTTTCGACTATGACATCGCCTTG 2204
Db 2105 AAGCGCATCATCTCCACCCCTTCTTCAATGATTTCACTTTCGACTATGACATCGCCTTG 2164
QY 2205 CTGAGCTGTGAGAGTCTGGTGGTGCAGACCGTCTGTCGCGCCCATCTGCTGCTGCTGAT 2264
Db 2165 CTGAGCTGTGAGAGTCTGGTGGTGCAGACCGTCTGTCGCGCCCATCTGCTGCTGCTGAT 2224
QY 2265 GCTACCATCTGCTTCTCTCTGTCGCAAGGCCCATCTGCGGTGCAGAGCTGGGGGCACAAAA 2324
Db 2225 GCCTCCCATCTGCTTCTCTCTGTCGCGCAAGGCCCATCTGCGGTGCAGAGCTGGGGGCACAAAA 2284
QY 2325 GAGGAGGTACCGAGCGCTGATCCTCGAAGAGGTGAGTCCGTGTATCAACCCAGAC 2384
Db 2285 TATGAGGCACTGGCGCGCTGATCCTGCAAAAGGTGAGATCCCGCTCATCAACCCAGAC 2344
QY 2385 ACCTGTGAGAGCTCATGCGCGAGCAGATCAACCCAGATGATGTGTGGGTTCCTC 2444
Db 2345 ACCTGTGAGAGCTCATGCGCGAGCAGATCAACCCAGATGATGTGTGGGTTCCTC 2404
QY 2445 AGTGGGGTGTGAGCTCCTGCGAGGGTGAATCTGCTGGGGCCCTTGTCAAGCGCGGAGAAA 2504
Db 2405 AGCGCGCGGTGAGCTCCTGCGAGGGTGAATTCGCGGGGACCCCTGTCCAGCTGAGGCG 2464
QY 2505 GATGGGCGAATGTTCCAGGCTGTGTGTGAGCTGGGTGAGGCTGGCTCAGAGGAAC 2564
Db 2465 GATGGGCGAATGTTCCAGGCTGTGTGTGAGCTGGGTGAGGCTGGCTCAGAGGAAC 2524
QY 2565 AAGCCAGGCTGTATACAAAGGCTCCCTGTAGTTTGGGACTGGATCAAGAGCACTGGG 2624
Db 2525 AAGCCAGGCTGTATACAAAGGCTCCCTGTAGTTTGGGACTGGATCAAGAGCACTGGG 2584
QY 2625 GTATAGCAGATGAGACAGACAGCGGACCAAAACACCCACAGGGATGCCCGCATGACACA 2684
Db 2585 GTATAGGCGCGG---GCCACCCAAATGTGTACACCTGCGGGGCCACCCATCTGTCCACC 2641
QY 2685 CCTGGATACAGGAGAGGAACATGACGACATTTATGCTGTGGCTCTCCCCCCCCCAACACA 2744
Db 2642 CCAGTGTGACG---CCTGAGGCTGGAGCTGGACCGCTGACTGCACCGCGCC---CCAGA 2699
QY 2745 ACCCAGACTGTGAATGCTATCTTAGACTCAGAT 2780
Db 2700 ACATACACTGTGAATCAATCTCCAGGCTCCAAAT 2735

XX DT 24-FEB-2003 (first entry)
XX DE Human membrane-type serine protease, MTSPl DNA.
XX KW Human; serine protease 14; CVSP14; cancer; malignancy; breast; colon;
KW gene therapy; cytostatic; membrane-type serine protease; enzyme; MTSPl;
XX OS gene; ds.
XX OS Homo sapiens.
XX FH Key Location/Qualifiers
XX FT CDS 23..2590
XX FT /*tag= a
XX FT /product= "Human MTSPl protein"
XX WO200277263-A2.
XX PN 03-OCT-2002.
XX PD 20-MAR-2002; 2002WO-US09039.
XX PF 22-MAR-2001; 2001US-278166P.
XX PR (CORV-) CORVAS INT INC.
XX PA Madison EL, Yeh J;
XX PI WPI; 2003-018938/01.
XX DR P-PSDB; AAE29791.
XX DT New purified CVSP14 polypeptide and encoding nucleic acid molecule,
XX PT useful for diagnosing, preventing and/or treating disorders, such as
XX PT cancers and malignancies of the breast, cervix, prostate, lung, ovary
XX PT or colon -
XX PS Disclosure; Page 167-171; 185pp; English.
XX CC The invention relates to transmembrane serine protease 14 (CVSP14), its
XX CC nucleic acid sequence and the method based on them. The methods and
XX CC compositions of the invention are useful for diagnosing, preventing
XX CC and/or treating conditions associated with the aberrant expression or
XX CC activity of the CVSP14 polypeptide, such as cancers and malignancies of
XX CC the breast, cervix, prostate, lung, ovary or colon. The methods are
XX CC also useful for identifying compounds that will modulate the protease
XX CC activity of CVSP14 polypeptide, and monitoring tumour progression and/or
XX CC therapeutic effectiveness. CVSP14 DNA used in gene therapy. The present
XX CC sequence is human membrane-type serine protease, MTSPl DNA.
XX SQ Sequence 3147 BP; 654 A; 952 C; 958 G; 583 T; 0 other;

Query Match 60.6%; Score 1883.2; DB 25; Length 3147;
Best Local Similarity 81.2%; Pred. No. 0;
Matches 2223; Conservative 0; Mismatches 508; Indels 5; Gaps 3;

QY 45 GATCGGACCCCAAAACCATGGGTAGCAATCGGGCCGCAAGCGCGAGGGGCTCTCAG 104
DB 5 GAGCGGCTCGGGTACCATGGGAGGCGATCGGCCCGCAAGGCGGAGGGGGCCCGAAG 64
QY 105 GACTTCGGCGGGGACTCAAGTACAACTCCCGGCTAGAGAACATGAATGGCTTTGAGGAG 164
DB 65 GACTTCGGCGGGGACTCAAGTACAACTCCCGGCAAGAAAGTGAATGGCTTTGAGGAA 124
QY 165 GGTGTGGAGTTCCTGCTCGGACAAATGCCAAGAAAGTGGAGAGCGAGGCCCGCAGCGC 224
DB 125 GGGTGGAGTTCCTGCGCAGTCAACAAAGTCAAGAAAGTGGAAAGCATGGCCGGGCGC 184
QY 225 TGGTGTGTGTGGGCGAGTGTCTTCAGCTTCCTCTTGTCTCTCCCTCATGGCTGGCTTG 284
DB 185 TGGTGTGTGTGGGCGGCTGTGATCGGCTCTCTTGTGTGTGTGGGATCGGCTTC 244
QY 285 CTGGTGTGGCACTTCCATTATCGGAATGTGGGTTTCAAAAAGTCTTCAATGGCCATCTG 344
|||||

DB 245 CTGGTGTGGCACTTTGCAGTACCGGGACGTGCGTGTCCAGAAGGTCTTCAATGGCTACATG 304
QY 345 AGGATCACAAATGAGATCTTTCTGGATGCGTATGAGAACTCCACCTCCACAGAGTTTATC 404
DB 305 AGGATCACAAATGAGAAATTTTGTGGATGCTTACGAGAACTCCAACTCCACTGATTTGTA 364
QY 405 AGCCTGGCAGCGCAGGTGAAGGAGCGCTGAAGCTCTGTGTACAAATGAAGTCCCTGTCTG 464
DB 365 AGCCTGGCCAGCAAGGTGAAGGACGCGCTGAAGCTGTGTACAGCGGAGTCCCATTCCTG 424
QY 465 GGTCCCTTACCACAAGAAGTCCGCTGTAAGTCCCTTCCAGTGAGGGCAGTGTCAATCGCTAC 524
DB 425 GGCCCTTACACAGGAGTCCGCTGTGACGCGCTTCCAGGAGGCGAGCGTCAATCGCTAC 484
QY 525 TACTGTCTAGAGTTTCCATCTCCCTCCACACCTTGGCAGAAAGAGTTTGAATCGCGCATGGCT 584
DB 485 TACTGTCTAGTTTCCAGCATCCCGCAGCACCTGTGTGAGGAGGCCCGAGCGCTCATGGCC 544
QY 585 GTGGAGCGAGTTGTAAACATTTGCCACCCCGAGACACGGGCACTGAAATCTTCTGTGTGTAACA 644
DB 545 GAGGAGCGGTAGTCAATGTCTGCCCGCGCGCGCTCTCCCTGAAGTCTTGTGTGTGCTCAC 604
QY 645 TCTGTGTGGCTTCCCTCATTTGACCCCGAGAAATGCTGTGAGAGGACTCAGGACACAGCTGC 704
DB 605 TCAGTGTGGCTTTCCCGACGACTTCAAAAACAGTACAGAGGACCCAGGACACAGCTGC 664
QY 705 AGTTTGGCTTCATGCCCATGGTGCAGCAGTGACACGCTTCTACTACCTTGGCTTCTCCC 764
DB 665 AGCTTTGGCTTGCACGCGCGGTGTGTGAGTGTATGCGCTTCCACACGCGCGGCTTCCCT 724
QY 765 ACAGTTCCTTACCGGGGCGATGCCGCTGCCAGTGGTCTTGGGGGAGACCGGACTCT 824
DB 725 GACAGCGCTTACCGCTCATGCCCTGCGAGTGGGCGCTTGGGGGAGACCGGACTCA 784
QY 825 GTCTGAGCTTCACTTCCGAGCTTTGATGTGCTGCTGCTGTGATGAGCATGCGAGTGC 884
DB 785 GTCTGAGCTTCACTTCCGAGCTTTGACCTTGTGCTTGTGCGAGCGCGCGCAGCGAC 844
QY 885 CTGGTCACTGTATGATAGCTTACCTGAGCCCGCATGGAAACCCACGCTGTGTGGCGGTGTGT 944
DB 845 CTGGTCACTGTATGATAGCTTACCTGAGCCCGCATGGAGCCCGCATGGAGCTGTGTGTGT 904
QY 945 GGCACCTTCTACCTTCTTACACCTGACTTCTCTCTTCTCTCCAGAACCTTCTCTTCTG 1004
DB 905 GGCACCTTCTCTCTCTTCTTACACCTGACTTCTCTCTCTCCAGAACCTTCTCTTCTG 964
QY 1005 ACCTGATTAACCAATAGTACCGCGGCGACATCTGCTGCTTGTGAGGCCACTTCTTCCAGCTG 1064
DB 965 ACCTGATTAACCAACACTGAGCGCGCGCATCCCGGCTTTGAGGCCACTTCTTCCAGCTG 1024
QY 1065 CCAAGATGAGCAGCTGTGGCGCTTTTGTGAGTGACACCCCAAGGGACATTTAGCAGCCCC 1124
DB 1025 CTTAGATGAGCAGCTGTGGAGCGCTTTACGTAAGCCCGAGGGGACATTTCAACAGCCCC 1084
QY 1125 TACTATCCAGGCGCACTACCGCGCCCAACATCAATGCACATGGAATCAAGTGGCCCAAC 1184
DB 1085 TACTATCCAGGCGCACTACCGCGCCCAACATCAATGCACATGGAATCAAGTGGCCCAAC 1144
QY 1185 AACCGGAACGTGAAGTGGCTTCAAACTCTTCTATCTGTGTGAGACCCCAACCTTACAGTG 1244
DB 1145 AACCGCATGTGAAGTGGAGTTCAAATCTTCTTCTTCTGAGGCGCGCGCTGCTGCG 1204
QY 1245 GGTCTCTGACCGAGACTATGTGGAGATCAACGGGGAGAAAGTACTGCGGTGAGAGTCC 1304
DB 1205 GGCACCTGCGCCCAAGGACTTACGTTGGAGATCAATGGGGAGAAATACTGCGGAGAGGCTCC 1264
QY 1305 CAGTTTGTGTGAGCAGCAACAGCAGCAAGATTTACAGTCCACTTCCATTTCTGATCACTCG 1364
DB 1265 CAGTTTGTGTGAGCAGCAACAGCAGCAAGATTTACAGTCCACTTCCATTTCTGATCACTCG 1324
QY 1365 TACACCGACACCGGGTCTCTAGCTAGTACCTCTCTTACGACTTCCAAACGACCGCTGCCCA 1424
DB 1325 TACACCGACACCGGGTCTCTAGCTAGTACCTCTCTTACGACTTCCAAACGACCGCTGCCCA 1384
|||||

1425 GGGATGTTTCATGTGCAAGACTGGAACGGTGCATCCGAAAGGAACTGCGCTGCGACGGCTGG 1484
1435 GGGAGTTTCAGTTCGCGCAGCGGGCGGTGTATCCGGAAGAGAGTGGCTGTGATGCTGG 1444
1485 GCAGATGCGCGGATATATAGTATGATGACGGTTACTGCGGATGCAATGCCACCCACCACTGTC 1544
1445 GCCGATGTCACCGACACACAGCGATGAGTCAACTGCGATTTGCGACGCGGCCACCACTGTC 1504
1545 ACCTGCAAAACACAGTTCCTGACAGCCCTCTCTGGGTCTGTGACAGTGTCAACGACTGT 1604
1505 ACCTGCAAGAACAGTTCCTGCAAGCCCTCTCTGGGTCTGTGACAGTGTGAAAGCACTGC 1564
1605 GGGGACCGGAAGTGACAGAGGAGGCTGCGAGTGTCTGCTGGGAGTTTCAAGTGTTCCAAT 1664
1565 GGAGACAAACAGCAGCGAGGGGTGAGTTGTCCGGCCACAGACCTTCAGTGTTCAT 1624
1665 GGGAGTGTCTCCCTCAGAGCCAGAGAGTGTAAATGGGAAGGACAACTGTGAGATGGGTCT 1724
1625 GGGAGTGTCTCTCGAAAGCCAGAGTGTCAATGGGAAGGACGACTGTGGGACGGGTCC 1684
1725 GAGAGGCTTCATGTGACAGCGTGAATGTCGTCTCTTGACACCAATATACCTACCGCTGC 1784
1685 GAGAGGCTTCCTGCCCAAGGTGAACGTCTGCTTGTACCAAAACACCTTACCCTGCTGC 1744
1785 CAAATGGCCCTCTGTGTGACAAAGGGCAACCTCTGAGTGTGATGGGAAGACGGACTGTAGC 1844
1745 CTCAATGGGCTCTGTGTGACAAAGGCAACCTCTGAGTGTGACGGGAAGGAGACTGTAGC 1804
1845 GATGGCTCCGATGAGAAAACTGTGACTGTGGCTGGATCCTTTTACAAACAGGCTGCG 1904
1805 GACGGCTCAGATGAGAAGGACTCGACTGTGGCTGCGGTCAATTCACGACACAGGCTCGT 1864
1905 GTGGTGTGGACGAAATGCGACGGGAGTGGCCCTGGCAGTGGACCTCCAGCC 1964
1865 GTTGTGGGGACGAGATGGGATGAGGGGAGTGGCCCTGGCAGGTAAAGCTGCGATGCT 1924
1965 CTGGGACAGGGCCACTTGTGTGGGGCTCGCTCATCTCTCTGACTGGCTGTCTCTGCA 2024
1925 CTGGGACAGGGCCACATCTGGGTGTCTCCCTCATCTCTCCCACTGGCTGTCTGCG 1984
2025 GCTCATGTCTTCAGGATGACAAATTTCAAGTACTCAGACTACAGATGTGAGCGGCC 2084
1985 GCACATGTCTACATCATGATGACAGAGGATTCAGGTACTCAGACCCCAACGCGAGCGGCC 2044
2085 TTCTGGGTCTGTGGACACAGCAAGCGAGTGGCTCTGGGGTGCAGAGTGAAGCTC 2144
2045 TTCTGGGTCTGCACACACAGCCAGCGCGCCCTGGGGTGCAGGAGCGAGGCTC 2104
2145 AAACGTATCATCACCCCTTCTTCAATGATTTCACTTCGACTATGACATCGCCTTG 2204
2105 AAGCGCATATCTCCACCCCTTCTTCAATGACTTCACCTTCGACTATGACATCGCGCTG 2164
2205 CTGAGCTGGAGAGTGGTGGAGTACAGACCGTGTGGCGCCCATCTGCTGCTGAT 2264
2165 CTGGAGTGGAGAAACCGGAGAGTACAGCTCCATGGTGGCGCCCATCTGCTGCGCGAC 2224
2265 GCTACCATGCTCTTCTGCTGGCAAGGCCATCTGGGTACAGGCTGGGGGCACAAAA 2324
2225 GCCTCCCATGCTTCTCTGCGGCAAGGCCATCTGGGTACCGGGCTGGGACACACCCAG 2284
2325 GAGGAGGTACCGAGCGCTGATCTCTGCAAGAGGTGAGATCCGTGTTCATCAACCAAGCC 2384
2285 TATGGAGGCACTGGCGGCTGATCTCTGCAAAAGGATGAGATCCGCGTCAATCAACCAAGCC 2344
2385 ACCTGTGAGGACCTCATGCGGAGAGATCACCCACAGATGATGTGTGGTTCCTC 2444
2345 ACCTGAGAGAACTCTCTGCGGAGAGATCACGCGCGCATGATGTGCGTTCCTC 2404
2445 AGTGGGGTGTGAGTCTCTGCGAGGGTGAATCTGTGGTGGCCCTTGTCAAGCGGAGAAA 2504
2405 AGCGGCGGCTGGACTCTCTGCGAGGGTGAATCTGCGGGGACCCCTGTCCAGCTGGAGGGC 2464

2505 GATGGCGGAATGTTCCAGGCTGGTGTGGTGAAGGCTGGCTCAGAGAAC 2564
2465 GATGGCGGATCTTCCAGGCGGTGTGGTGAAGCTGGGAGACGGCTGGCTCAGAGAAC 2524
2565 AAGCCAGGCGGTGTACAAAGGCTCCCTGTAGTTTGGGACTGGATCAAAGAGCACTGGG 2624
2525 AAGCCAGGCGGTGTACAAAGGCTCCCTGTAGTTTGGGACTGGATCAAAGAGAACTGGG 2584
2625 GTATAGCAGCATGGACAGACGCGGACCAACACCCACAGGGATGCCGACATGCACA 2684
2585 GTATAGGGGCCGGG---GCCACCCAAATGTGTACACTGCGGGGCCACCCATCGTCCACC 2641
2685 CTTGGATACAGGAGAGAAACACTGACGACATTTATGCTGTGGCTCTCCCGCCCAACACA 2744
2642 CCAGTGTGACG-CCTGACAGGCTGGAGACTGGACCGTGAATGCACAGGCGCC-CCAGA 2699
2745 ACCAGACTGTGAACCTGCAATCCCTTAGGACTCAGAT 2780
2700 ACATACACTGTGAATCTCAATCTCCAGGGCTCCAAAT 2735

Search completed: February 21, 2004, 02:22:12
Job time : 812 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 21, 2004, 02:07:32 ; Search time 179 Seconds
(without alignments)
7658.860 Million cell updates/sec

Title: US-09-900-751-1

Perfect score: 3106

Sequence: 1 catggtgacggtgccggg.....ttaaaaaaaaaaaaaaaaaa 3106

Scoring table: IDENTITY_NUC

Gapop.10.0 , Gapext 1.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents NA.*

- 1: /cgn2_6/ptodata/2/ina/5A COMB.seq.*
- 2: /cgn2_6/ptodata/2/ina/5B COMB.seq.*
- 3: /cgn2_6/ptodata/2/ina/6A COMB.seq.*
- 4: /cgn2_6/ptodata/2/ina/6B COMB.seq.*
- 5: /cgn2_6/ptodata/2/ina/PCTUS COMB.seq.*
- 6: /cgn2_6/ptodata/2/ina/backfiles.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1883.2	60.6	3147	2	US-09-027-337-1
2	1883.2	60.6	3147	4	US-09-644-600-1
3	1883.2	60.6	3147	4	US-09-644-600-18
4	1530.8	49.3	2900	2	US-09-027-337-9
5	1530.8	49.3	2900	4	US-09-644-600-9
6	701.2	22.6	1553	4	US-09-280-116-10
7	306.8	9.9	434	4	US-09-702-705-1480
8	306.8	9.9	434	4	US-09-736-457-1480
9	173.2	5.6	796	4	US-09-280-116-107
10	113.6	3.7	2413	3	US-09-518-046-1
11	109.8	3.5	2544	3	US-09-342-749-3
12	101.6	3.3	1479	3	US-09-342-749-1
13	101.6	3.3	1479	4	US-09-691-840-1
14	101.2	3.3	1128	2	US-09-016-366A-20
15	101.2	3.3	1128	2	US-09-016-366A-15
16	100.6	3.2	1077	3	US-08-807-151-2
17	100.6	3.2	1077	3	US-08-807-151-2
18	99.6	3.2	1081	2	US-09-016-366A-22
19	99.6	3.2	1081	2	US-08-978-404B-17
20	98.4	3.2	2479	3	US-08-342-749-29
21	98.4	3.2	2479	3	US-09-691-840-29
22	98.2	3.2	735	3	US-09-079-970A-1
23	98.2	3.2	771	3	US-09-079-970A-4
24	98	3.2	1137	2	US-09-016-366A-18
25	98	3.2	1137	2	US-08-978-404B-13
26	94.2	3.0	2416	3	US-08-261-416-1
27	92.2	3.0	1605	2	US-09-000-846-1

28	91	2.9	980	4	US-09-023-942A-30	Sequence 30, Appli
29	91	2.9	1110	4	US-09-386-653A-1	Sequence 1, Appli
30	90.8	2.9	901	1	US-08-508-448C-9	Sequence 9, Appli
31	90.8	2.9	1460	4	US-09-370-838-80	Sequence 80, Appli
32	90.8	2.9	1462	4	US-09-370-838-55	Sequence 55, Appli
33	90.8	2.9	1517	1	US-08-508-448C-15	Sequence 15, Appli
34	90.8	2.9	2790	4	US-09-370-838-79	Sequence 79, Appli
35	90.4	2.9	1783	3	US-09-510-738A-188	Sequence 188, App
36	90.4	2.9	1783	4	US-09-861-966-188	Sequence 188, App
37	90.4	2.9	2363	4	US-09-742-703-3	Sequence 3, Appli
38	89.4	2.9	1154	2	US-09-016-366A-16	Sequence 16, Appli
39	89.4	2.9	1154	2	US-08-978-404B-11	Sequence 11, Appli
40	89.4	2.9	1212	4	US-09-620-312D-431	Sequence 431, App
41	88.8	2.9	1130	4	US-09-386-653A-8	Sequence 8, Appli
42	88.8	2.9	2038	3	US-09-008-271A-18	Sequence 18, Appli
43	88.8	2.9	2079	4	US-09-656-002-1	Sequence 1, Appli
44	88.6	2.9	1615	4	US-08-820-002-1	Sequence 1, Appli
45	88.4	2.8	696	1	US-08-508-448C-24	Sequence 24, Appli

ALIGNMENTS

RESULT 1

US-09-027-337-1

; Sequence 1, Application US/09027337B

; Patent No. 5972616

; GENERAL INFORMATION:

; APPLICANT: O'Brien, Timothy J.

; APPLICANT: Tanimoto, Hirotooshi

; TITLE OF INVENTION: TADG-15: An Extracellular Serine Protease Overexpressed in

; TITLE OF INVENTION: Breast and Ovarian Carcinomas

; FILE REFERENCE: D6064

; CURRENT APPLICATION NUMBER: US/09/027,337B

; CURRENT FILING DATE: 1998-02-20

; SEQ ID NO 1

; LENGTH: 3147

; TYPE: DNA

; ORGANISM: Homo sapiens

; FEATURE:

; LOCATION: 23..2589

; OTHER INFORMATION: cDNA sequence of TADG-15

US-09-027-337-1

Query Match	60.68;	Score	1883.2;	DB 2;	Length	3147;
Best Local Similarity	81.28;	Pred. No. 0;				
Matches 2223;	Conservative	0;	Mismatches	508;	Indels	5;
Gaps	3;					
Qy	45	GATCGGACCGCCAAACCATGGGTAGCAATCGGGCGCGCAAGCGCGGAGGGGGCTCTCAG	104			
Db	5	GAGCGGCTCGGGGTACCATGGGAGCGATCGGGCGCGCAAGCGCGGAGGGGGCCCGAAG	64			
Qy	105	GACTTGGCGCGGAGCTCAAGTCAACTCCGGCTAGAGAACATGAATGGCTTTAGAGAG	164			
Db	65	GACTTGGCGCGGAGCTCAAGTCAACTCCGGCGCAGAGAAAGTGAATGGCTTTAGAGAGAA	124			
Qy	165	GGTGTGGAGTTCTGCTCGCGCAATGCCAAGAAAGTGGAGAGCGAGGCCCGCCAGCGC	224			
Db	125	GGCGTGGAGTTCTGCGCAGTCAACACGTCAAGAGGTGAAAGCATGGCCCGGGCGC	184			
Qy	225	TGGGTGGTGTGGTGGCAGTGTTCAGTTCCTCTTGTCTCTCCCTCATGTGGTGGCTTG	284			
Db	185	TGGGTGGTGTGGCAGCGCGTGTGATCGGCTCTCTTGTCTTGTGGGATCGGCTTC	244			
Qy	285	CTGGTGGCATTTCATTATCGAATGTGCGGGTCAAAAAGTCTTCAATGCCATCTG	344			
Db	245	CTGGTGGCATTTCAGTACCGGACGTCGCTGTCCAGAAAGTCTTCAATGGCTACATG	304			
Qy	345	AGGATCACAAATCAGATCTTCTGGATGCGTATGAGACTCCACCTCCACAGAGTTTATC	404			
Db	305	AGGATCACAAATGAGAATTTTGTGGATGCGTACGAGAACTCCAACTCCACTGAGTTTGA	364			

405 AGCTGGCCAGCCAGGTGAAGGAGCGCTGAAGCTGCTGTACAAATGAAGTCCCTGTCTCTG 464
Db
365 AGCTGGCCAGCCAGGTGAAGGAGCGCTGAAGCTGCTGTACAGCGAGTCCATTCCTG 424
QY
465 GGTCCCTTACACAAAGAGTGGCTGTAACTGCTTCACTGAGGAGAGTGTATCGCTTAC 524
Db
425 GGCCTTACACAAAGAGTGGCTGTAACTGCTTCACTGAGGAGAGTGTATCGCTTAC 484
QY
525 TACTGTCTAGAGTTAGCATCCCTCCACACCTGGCAGAGAGAGTTGATCGGCCATGGCT 584
Db
485 TACTGTCTAGAGTTAGCATCCCTCCACACCTGGCAGAGAGAGTTGATCGGCCATGGCT 544
QY
585 GTGGAGCGAGTTGTAACTTGGCCACCCCGAGAGCGGCACTGAATCTCTTGTGCTTAA 644
Db
545 GAGGAGCGGAGTGTATGCTGTGCTCCCGCGCGCTCCCTGAAGTCTTGTGTGCTCACC 604
QY
645 TCTGTGTGGCTTCCCTTACCTTGAATGCTCCAGAGTCTCCAGAGAGTGTATCGCTTAC 704
Db
605 TCACTGTGTGGCTTCCCTTACCTTGAATGCTCCAGAGTGTATCGCTTAC 664
QY
705 AGTTTGGCTGTATGCTGTGCTGAGAGTGTATGCTTCACTACCTTGGCTTCCCT 764
Db
665 AGTTTGGCTGTATGCTGTGCTGAGAGTGTATGCTTCACTACCTTGGCTTCCCT 724
QY
765 AACAGTCCCTTACCTTGGCTGTATGCTGTGCTGAGAGTGTATGCTTCACTACCTTGGCTTCCCT 824
Db
725 GACAGGCTTACCTTGGCTGTATGCTGTGCTGAGAGTGTATGCTTCACTACCTTGGCTTCCCT 784
QY
825 GTGCTGAGCTTACCTTGGCTGTATGCTGTGCTGAGAGTGTATGCTTCACTACCTTGGCTTCCCT 884
Db
785 GTGCTGAGCTTACCTTGGCTGTATGCTGTGCTGAGAGTGTATGCTTCACTACCTTGGCTTCCCT 844
QY
885 CTGGTCACTGCTGTATGCTGTGCTGAGAGTGTATGCTTCACTACCTTGGCTTCCCT 944
Db
845 CTGGTCACTGCTGTATGCTGTGCTGAGAGTGTATGCTTCACTACCTTGGCTTCCCT 904
QY
945 GGCACCTTCTACCTTGGCTGTATGCTGTGCTGAGAGTGTATGCTTCACTACCTTGGCTTCCCT 1004
Db
905 GGCACCTTCTACCTTGGCTGTATGCTGTGCTGAGAGTGTATGCTTCACTACCTTGGCTTCCCT 964
QY
1005 AGCTGATACCAATACTGACCGGCGACATCTGGCTTTGAGGCGACATTTAGCGAGCTG 1064
Db
965 AACTGATACCAATACTGACCGGCGACATCTGGCTTTGAGGCGACATTTAGCGAGCTG 1024
QY
1065 CCAAGATGAGCAGCTGTGGCGCTTTTGAAGTGAACCCCAAGGAGACATTTAGCGAGCTG 1124
Db
1025 CCAAGATGAGCAGCTGTGGCGCTTTTGAAGTGAACCCCAAGGAGACATTTAGCGAGCTG 1084
QY
1125 TACTATCAGGCGCACTACCGGCGCAATCACTGACATGGAATATCAAGTGGCGCAAC 1184
Db
1085 TACTATCAGGCGCACTACCGGCGCAATCACTGACATGGAATATCAAGTGGCGCAAC 1144
QY
1185 AACCGGAACTGAGGTGGCTTCAAACTTCTATCTGCTGAGCCCAAGCTGACAGTG 1244
Db
1145 AACCGAACTGAGGTGGCTTCAAACTTCTATCTGCTGAGCCCAAGCTGACAGTG 1204
QY
1245 GGTCTCTGACCAAGGACTATGTGAGATCAACGGGAGAACTATGCTGAGAGCTG 1304
Db
1205 GGTCTCTGACCAAGGACTATGTGAGATCAACGGGAGAACTATGCTGAGAGCTG 1264
QY
1305 CAGTTTGTGGTGAAGCAACAGCAGCAGATGTACAGTCACTTCACTGATCTGATCTG 1364
Db
1265 CAGTTTGTGGTGAAGCAACAGCAGCAGATGTACAGTCACTTCACTGATCTGATCTG 1324
QY
1365 TACAGGCAACCGGTTCTTGTGAGTGTCTCTGAGTGTCTTCACTGAGTGTCTTCACTGAGTGTCT 1424
Db
1325 TACAGGCAACCGGTTCTTGTGAGTGTCTCTGAGTGTCTTCACTGAGTGTCTTCACTGAGTGTCT 1384
QY
1425 GGTGATTTGATGCAAGACTGGAAGGCTGATCGGAGCAACCTGCTGCGAGCGCTGG 1484
Db
1385 GGTGATTTGATGCAAGACTGGAAGGCTGATCGGAGCAACCTGCTGCGAGCGCTGG 1444
QY
1485 GCAGACTGCGCGGATTTAGTGTAGAGCTTACTGCGGATGCAATGCCACCCAGCTTC 1544

1445 GCCGACTGACCGACCAACGCGATGAGCTCAACTGAGTTGGACCGCCGCGCACAGATTTC 1504
QY
1545 ACGTCAAAAAACAGTTCTGCAAGCCCTCTTCTGGTCTGTGACAGTGTCAACGACTGT 1604
Db
1505 ACGTCAAAAAACAGTTCTGCAAGCCCTCTTCTGGTCTGTGACAGTGTGAACGACTGT 1564
QY
1605 GGGGACGGAAGTGAAGGAGGCTGAGCTGCTGCTGGGAGTTTCAAGTGTTCAT 1664
Db
1565 GAGACAACAGCGAGAGAGGGGTGAGTTGTCGGGCCACAGCTTCAAGTGTTCAT 1624
QY
1665 GGGAAAGTGTCTCCCTCAGAGCCAGAGTGTAAATGGGAAGCAAACTGTGAGAGTGGTCT 1724
Db
1625 GGGAAAGTGTCTCCCTCAGAGCCAGAGTGTAAATGGGAAGCAAACTGTGAGAGTGGTCT 1684
QY
1725 GAGAGGCTTCAATGTGACAGCTGAAATGTGCTTCTTGACCAAAATATACCTTACCGCTGC 1784
Db
1685 GACGAGGCTTCTGCTCCCAAGGTGAACGCTGCTCACTTGTACCAAAACACACCTTACCGCTGC 1744
QY
1785 CAAAATGGCTTCTGCTGAGCAAGGCGAACCTGAGTGTGATGGGAAGCAAGCTGTAGC 1844
Db
1745 CTCATGGCTTCTGCTGAGCAAGGCGAACCTGAGTGTGACGGGAAGGAGCTGTAGC 1804
QY
1845 GATGCTTCGATGAGAAAACTGTGCTGAGCTGCGCTGCGATCTCTTACCAAAACAGGCTGC 1904
Db
1805 GACGCTCAGATGAGAGGACTGCGACTGTGGCTGCGCTCACTTACGAGACAGGCTGT 1864
QY
1905 GTGGTGTGGGCAAAATGCGGACGAGGCGAGTGGCTTGGCAGGTGAGCTTCAACGCT 1964
Db
1865 GTGGTGTGGGCAAAATGCGGACGAGTGGCTTGGCAGGTGAGCTTCAACGCT 1924
QY
1965 CTGGGCGAGGGCACTTGTGGGCGCTGCTCACTCTCTCTGACTGGCTGCTCTGCA 2024
Db
1925 CTGGGCGAGGGCACTTGTGGGCGCTGCTCACTCTCTCTGACTGGCTGCTCTGCA 1984
QY
2025 GCTCATGTCTTTCAGATGACAAAAATTTCAAGTACTCAGACTACACGATGTGGAGCGGC 2084
Db
1985 GCACACTGTATCATCATGACAGAGGATTCAGGTACTCAGACCCCAAGCAGTGGAGCGGC 2044
QY
2085 TTCCTGGGCTGCTGCGACGAGGAGCGAGTGGCTCTGGGGTGCAGAGCTGAAGCTC 2144
Db
2045 TTCCTGGGCTGCGACGAGGAGCGAGTGGCTCTGGGGTGCAGAGCTGAAGCTC 2104
QY
2145 AAACCTATCATCAACCAACCTTCTCAATGATTTTCACTTCACTATGACATCGCTTG 2204
Db
2105 AAGCGCATCATCTCCACCTTCTTCAATGATTTTCACTTCACTATGACATCGCTTG 2164
QY
2205 CTGGAGCTGGAGAGTGGTGGAGTACAGACCGTGTGGCGCCCATCTGCTGCTGTAT 2264
Db
2165 CTGGAGCTGGAGAGTGGTGGAGTACAGACCGTGTGGCGCCCATCTGCTGCTGTAT 2224
QY
2265 GCTACCCATGTCTTCTTCTGCGCAAGGCGCATCTGGGTCAACAGGCTGGGGGCACACAAA 2324
Db
2225 GCCTCCCATGTCTTCTTCTGCGCAAGGCGCATCTGGGTCAACAGGCTGGGGGCACACAAA 2284
QY
2325 GAGGAGGTACCGGAGCGCTGATCTTCTGAGAGGCTGAGATCCGTGTCTATCAACAGAGCT 2384
Db
2285 TATGAGGCACTTGGCGCTGATCTTCTGCAAAAGGCTGAGATCCGTGTCTATCAACAGAGCT 2344
QY
2385 ACCTGTGAGGAGCTCATGCGCGAGAGATTCACCCCAAGATGTGTGTGGTTCCTC 2444
Db
2345 ACCTGTGAGGAGCTCATGCGCGAGAGATTCACCCCAAGATGTGTGTGGTTCCTC 2404
QY
2445 AGTGGGCTGTGGACTCTGCTGCGAGGCTGCTTGTGTGGCTTCTCAAGCGCGAGAAA 2504
Db
2405 AGCGGCGGCTGGATCTCTGCGAGGCTGATTCGCGGAGACCTCTGTCCAGGTGGAGGG 2464
QY
2505 GATGGCGGAATTTTCCAGGCTGGTGTGGTGGAGTGGGCTGAAGGCTGCGCTCAGAGGAA 2564
Db
2465 GATGGCGGATTTTCCAGGCTGGTGTGGTGGAGTGGGCTGCGCTCAGAGGAA 2524
QY
2565 AAGCCAGGCTGTACACAGGCTTCTGTAGTTGGGACTGGATCAAGAGGACACTGGG 2624

Db 2525 AAGCCAGGCGGTACACAAAGGCTCCCTCTGTTTCGGGACTGGATCAAAAGAGAACACTGGG 2584
Qy 2625 GTATAGCAGCATGACAGACAGCGGACCAACACACACAGGATGCCGACATGACACA 2684
Db 2585 GTATAGGGGCGGG--GCCACCAATATGTATACCTTGGGGGCCACCATGTCTCCACC 2641
Qy 2685 CTTGGATACAGGAGAGGAACAACATGACACATTTATGTGTGGCTCCCTCCCTCCCAACACA 2744
Db 2642 CCAAGTGTGACG-CCTGCAGSCTGGAGACTGGACCGTGAAGTCAACACAGCGGCC-CCAGA 2699
Qy 2745 ACCAGACTGTGAAGTCAATCTTAGACTCAGAGT 2780
Db 2700 ACATACACTGTGAAGTCAATCTCCAGGGCTCCAAAT 2735

RESULT 2
US-09-644-600-1
; Sequence 1, Application US/09644600
; Patent No. 6451500
; GENERAL INFORMATION:
; APPLICANT: O'Brien, Timothy J.
; APPLICANT: Tanimoto, Hiroto
; TITLE OF INVENTION: TADG-15: An Extracellular Serine Protease
; TITLE OF INVENTION: Overexpressed in Carcinomas
; FILE REFERENCE: D6064CIP/D
; CURRENT APPLICATION NUMBER: US/09/644,600
; PRIOR FILING DATE: 2000-08-23
; PRIOR APPLICATION NUMBER: 09/421,213
; PRIOR FILING DATE: 1999-10-20
; PRIOR APPLICATION NUMBER: 09/027,337
; PRIOR FILING DATE: 1998-02-20
; NUMBER OF SEQ ID NOS: 98
; SEQ ID NO 1
; LENGTH: 3147
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: TADG-15
US-09-644-600-1

Query Match 60.6%; Score 1883.2; DB 4; Length 3147;
Best Local Similarity 81.2%; Pred. No. 0;
Matches 2223; Conservative 0; Mismatches 508; Indels 5; Gaps 3;

Qy 45 GATCGGACCGCCAAACCATGGGTAGCAATCGGGCGGCAAGCGCGGAGGGGGCTCTCAG 104
Db 5 GAGCGGCTCGGGGTACCATGGGAGCGATCGGCGCCGCAAGGCGGAGGGGGCCGAG 64
Qy 105 GACTTCGGCGGGGACTCAAGTACAATCCCGGCTAGAGAACATGAATGGCTTTGAGGAG 164
Db 65 GACTTCGGCGGGGACTCAAGTACAATCCCGGCAAGGAAAGTGAATGGCTTTGAGGAA 124
Qy 165 GGTGTGAGTTCTCGCTCGCAACAAAGTGGAGAGAGGAGGAGGCGCCAGGCGC 224
Db 125 GCGGTGAGTTCTCGCAAGTCAAGTGAAGGTTGGAAGAGCATGGCCCGGGGCGC 184
Qy 225 TGGGTGTGTGTGTGGAGTCTGTTCAGCTTCTCTTGTCTCTCCCTCATGGCTGGCTTG 284
Db 185 TGGGTGTGTGTGGAGCGGTGTGATCGGCTCTCTTGTGTGTGTGGGATCGGCTTC 244
Qy 285 CTGTGTGGGACTTCATATATGGAATGTGGGGTTCAAAAGTCTTCAATGGCCATCTG 344
Db 245 CTGTGTGGGACTTCGAGTACCGGACGTCGTGTCCAGAAAGTCTTCAATGGCTACATG 304
Qy 345 AGGATCAAAATGAGATCTTCTGATGCGTATGAGACTCCACTCCACAGAGTTTATC 404
Db 305 AGGATCAAAATGAGATTTTGTGATGCGTATGAGACTCCACTCCACTGAGTTTGTGA 364
Qy 405 AGCTTGGCGGAGGTTGAAGGAGCGCTGAAGTGTGTGATGATGAGTCCCTGTCTGT 464
Db 365 AGCTTGGCGGAGGTTGAAGGAGCGCTGAAGTGTGTGATGATGAGTCCCTGTCTGT 424
Qy 465 GGTCCCTACCAAGAGTGTGGTGTAACTGCTTTCAGTGGGAGGAGTGTATCGCTTAC 524

Db 425 GGCCCTTACCAAGAGTGGCTGTGAGCGCTTTCAGCGAGGCGACGCTCATCGCTAC 484
Qy 525 TACTGTGAGATTGAGATCCCGCCACACTGCGCAGAGAGTTGATGGGCCATGGCT 584
Db 485 TACTGTGCTGAGTTGAGCATCCCGCAGACCTTGGTGGAGGAGCCGAGCGCTCATGCC 544
Qy 585 GTGAGCGAGTTGTAACATTCGCCACCCGAGCAGCGGCACTGAAATCTTTCGTCTAAACA 644
Db 545 GAGAGCGGCTAGTCTGCTGCTCCCGCGGGCGCTCCCTGAAGTCTTGTGTGCTCACC 604
Qy 645 TCTGTGTGTGGCTTCCCGCATTTGACCCCGAGAAATGCTGACAGGACTCAGACAAACAGTGC 704
Db 605 TCAGTGGTGGCTTTTCCCGCAGACTCCAAAACAGTACAGAGGACCCAGAGCAACAGCTGC 664
Qy 705 AGTTTTGCTGCTGATGCCCATGTTGAGCAGTACACGCTTTCACCTCCCTGGCTTCCCC 764
Db 665 AGCTTGTGGCTGACGCGCCCGGTGTGAGCTGATGCGCTTTCACACGCGCGGCTTCCCT 724
Qy 765 AACAGTCCCTTACCCGCGCATGCGCTGCGAGTGGGTCTCTGGGGGGGAGCGGACTCT 824
Db 725 GACAGCCCTTACCCCGCTCATGCGCTGCGAGTGGGCCCTGCGGGGGAGCGCGACTCA 784
Qy 825 GTGCTGAGCTCTACCTTCCGAAAGTGTGATGCTGCTCCCTGTGATGAGCATGGCAGTGAC 884
Db 785 GTGCTGAGCTCTACCTTCCGAGCTTTCAGCTTGGCTCTCTGCGCAGCAGCGCGCAGCAG 844
Qy 885 CTGCTCACCGTGTATGATAGCTGAGCCCATGGAACCCCGCAGCTGTGGTGGCGCTGTGT 944
Db 845 CTGCTGAGCTGTATCAACACCTTGAAGCCCATGAGCCCGCAGCCCTGGTGTGAGTGTGT 904
Qy 945 GGCACCTTCTCACCTCTCTAACACTGACTTTCCTCTCTCCAGAACGCTTCTCTTGTG 1004
Db 905 GGCACCTTACCTCTCTCTAACACTGACTTTCCTCTCTCCAGAACGCTTCTCTCATC 964
Qy 1005 AGCTGATTAACCAATGATCCGCGGACATCTCTGGCTTTGAGGCGCATTTCTTCCAGCTG 1064
Db 965 ACATGTATAACCAACACTGAGCGCGGCATCCCGGCTTTGAGGCGCATCTTCTTCCAGCTG 1024
Qy 1065 CCAAGATGAGCAGCTGTGGCGGCTTTTGTAGTGACACCAAGGAGCATTTAGCAGGCCCC 1124
Db 1025 CTTAGGATGAGCAGCTGTGGAGCGCGCTTACGTAAAGCCCGAGGAGCATTTCAACAGCCCC 1084
Qy 1125 TACTATCCAGGCGCACTACCGCCCAACATCAACTGCACTGCAATGGAATATCAAGGTGCCCCAAC 1184
Db 1085 TACTATCCAGGCGCACTACCGCCCAACATTTGACTGCACTGGAACATTTGAGGTGCCCCAAC 1144
Qy 1185 AACCGGAAGTGAAGTGGCTTCAAACTCTTCTATCTGTGTGGACCCCAAGCTACCACTG 1244
Db 1145 AACCGCATGTGAAGTGAAGTCTTCAAACTTCTTCTGTGTGGAGCGCGGCTGCGCTGG 1204
Qy 1245 GGCTCTCTGACCAAGGACTATGTGGAGTCAACCGGGGAGAGTACTGCGGTGAGAGGTCC 1304
Db 1205 GGCACTGCGCCCAAGGACTAGTGGAGTCAATGGGGGAAATACTGCGGAGAGAGGTCC 1264
Qy 1305 CAGTTGTGTGTGAGCAGCAACAGCAGCAAGTATACGTCCACTTCCATTTGATCATCTG 1364
Db 1265 CAGTTGTGTGTGAGCAGCAACAGCAGCAAGTATACGTTCGCTTCCACTCAGATCAGTCC 1324
Qy 1365 TACACGAGACCGGGTCTTCTAGCTGAGTACTCTCTACGACTCCACGACCGCGTCCCA 1424
Db 1325 TACACGAGACCGGGTCTTCTAGCTGAAATCTCTCTACGACTCCAGTACCGGCTGCGCG 1384
Qy 1425 GGGATGTTCTATGTGCAAGACTGGAACGCTGATCCGAAAGGAAATGCGCTGCGAGCGGTGG 1484
Db 1385 GGGCAGTTCACTGCGCCGACGCGGGCGGTGTATCCGGAAGAGCTGCGCTGTGATGGCTGG 1444
Qy 1485 GCAGATGCGCGGATTTATAGTGTAGCGGTATCTGCGGATGCAATGCCACCGCAGTTTC 1544
Db 1445 GCCGACTGCAACCGACACAGCGATGAGCTCAACTGCAAGTTTGGCAGCGCGGCGCACAGTTTC 1504
Qy 1545 AGTGTGAAAACAGTTCTGAGGCGGCTTCTTGGGTCTGTGACAGTGTCTACGACTGT 1604

Db 1505 ACCTGACGAAGTCTTCAAGCCCTCTCTGGGCTCTGCACAGTGTGAACGACTGC 1564
 QY 1605 GGGACGGAAGTGAACGAGGAGGCTGAGCTGCTCTCTGGGAGTTTCAAGTGTTCACAT 1664
 Db 1565 GGAGACACGACGACGAGGAGGCTGAGTGTTCGGGCCAGACCTTCAGTGTTCACAT 1624
 QY 1665 GGAAGTGTCTCCCTCAGAGCCAGAAAGTGTAAATGGGAAGACAACTGTGTGAGATGGGTCT 1724
 Db 1625 GGAAGTGTCTCTCGAAAGCCAGCAGTGTCAATGGGAAGACGACTGTGTGGGGAGCGGTCC 1684
 QY 1725 GACGAGGCTTCAGTGTGACGAGTGTGATGTCTCTGTGACCAAAATATACCTACCGCTGC 1784
 Db 1685 GACGAGGCTCTCTCCCAAGGTGAAGCTGTCACTTGTACCAAAACACACTACCGCTGC 1744
 QY 1785 CAAAATGGCTCTCTGTGACGAGGCAACCTCAGTGTGTGATGGGAAGAGAGCTGTAGC 1844
 Db 1745 CTCAATGGCTCTCTGTGACGAGGCAACCTCAGTGTGTGACGGAAGAGAGCTGTAGC 1804
 QY 1845 GATGGCTCCGATGAGAAAACTGTGACTGTGGGCTGGATCTCTTTTACCAACAGGCTGCG 1904
 Db 1805 GACGGCTCAGATGAGAAAGGACTGCACTGTGGGCTGGGCTCAATTCAGGACAGAGCTCGT 1864
 QY 1905 GTGGTGTGTCGACGATGCGGAGGAGGAGTGGCCCTGGCAGGTGAGCCTCCAGCC 1964
 Db 1865 GTTGTGGGGCAGCGATGCGATGAGGGCAGTGGCCCTGGCAGGTGAGCCTGCTGCT 1924
 QY 1965 CTGGCCAGGGCCACTTGTGTGGGGCCCTCGCTATCTCTCTGACTGGCTGTCTGCA 2024
 Db 1925 CTGGCCAGGGCCACTTGTGGGTGTCTCTCTATCTCTCCAACTGGCTGTCTGCTG 1984
 QY 2025 GCTCATTTCTTCAGGATGACAAAAATTTCAAGTACTCAGACTACAGATGTGAGCGGCC 2084
 Db 1985 GCACACTGCTACATGATGACGAGGATTCAGGTACTCAGACCCCAAGCAGTGGAGCGCC 2044
 QY 2085 TTCTGTGCTCTGTGGACACGAGCAAGCGAGTGCCTCTGGGGTGACGAGCTGAAGCTC 2144
 Db 2045 TTCTGTGGCTTGCAGACACGAGCAGCGAGCCCTCTGGGGTGACGAGCGCAGCTC 2104
 QY 2145 AAAGTATCATACCCACCTCTCTCAATGATTTCACTTCCGATGATGATGATGATGATG 2204
 Db 2105 AAGCGCATCATCTCCACCCCTCTTCAATGATCTTCACTTCCGATGATGATGATGATG 2164
 QY 2205 CTGGAGCTGAGAGTGTGGTGGATGACGACCGTCTGTGGCCCAATCTGCTGCTGCTGAT 2264
 Db 2165 CTGGAGCTGAGAAACCGGACAGTACAGTCTTCCATGTTGGTGGCCCAATCTGCTGCGGAC 2224
 QY 2265 GCTACCCATGTCTTCTCTGTGGCAAGGCCATCTGGGTTCACAGGCTGGGGGCAACAAAA 2324
 Db 2225 GCCTCCCATGTCTTCTCTGCGGCAAGGCCATCTGGGTTCACAGGCTGGGGGCAACCCAG 2284
 QY 2325 GAGGAGGTACCGAGCGCTGATCTCTGAGAGGCTGAGATCCGTGTATCAACACGAGCC 2384
 Db 2285 TATGAGGCACTGGCGGCTGATCTGCAAAAGGTGAGATCCGCGTCAACACGAGCC 2344
 QY 2385 ACCTGTGAGGACCTCATGCGCAGCATACCCACGAAATGATGTGTGGGTTCCTC 2444
 Db 2345 ACCTGTGAGAAACCTCTCTGCGCAGCATACCGCGCATGATGTGTGGGTTCCTC 2404
 QY 2445 AGTGGGGGTGTGAGTCTCTGCGAGGGTGAATCTGTGGTGGCCCTTGTCAAGCGCGGAGAA 2504
 Db 2405 AGCGGCGGTGAGTCTCTGCGAGGGTGAATCTGCGGGGACCCCTGTCCAGCGTGGAGCG 2464
 QY 2505 GATGGGGAATGTTCAGGCTGTGTGGTGTGAGTGGGTGAGGCTGTGCTGAGAGAAC 2564
 Db 2465 GATGGGGAATGTTCAGGCTGTGTGGTGTGAGTGGGTGAGGCTGTGCTGAGAGAAC 2524
 QY 2565 AAGCAGCGGTGTACAGAGCTCTCTGAGTGTGGTGTGAGTGGGTGAGGCTGTGCTGAGAG 2624
 Db 2525 AAGCAGCGGTGTACAGAGCTCTCTGAGTGTGGTGTGAGTGGGTGAGGCTGTGCTGAGAG 2584
 QY 2625 GTATAGCAGCATGACAGACAGCGACCAACCAACCAAGGATGCCGACATGCACATGCACA 2684
 Db 2585 GTATAGGCGCGGG---GCCACCAATGTGTACCTGTGGGGGCCACCCATGCTCCACC 2641

QY 2685 CTGGATACAGGAGAGAACTGACGACATTTATGCTGTGGCTCTCCCCCCCCAACACA 2744
 Db 2642 CCAGTGTGCACG-CTTGAGGCTGGAGACTGGACCGCTGACTGCACAGGCGCC-CCAGA 2699
 QY 2745 ACCGAGCTGTGAACTGCACTCTTAGGACTCAGAGT 2780
 Db 2700 ACATACACTGTGAACCTCAATCTCCAGGCTCCAAAT 2735

RESULT 3

US-09-644-600-18/c
 ; Sequence 18, Application US/09644600
 ; Patent No. 6451500
 ; GENERAL INFORMATION:
 ; APPLICANT: O'Brien, Timothy J.
 ; APPLICANT: Tanimoto, Hirotsoshi
 ; TITLE OF INVENTION: TADG-15: An Extracellular Serine Protease
 ; FILE REFERENCE: Overexpressed in Carcinomas
 ; CURRENT FILING DATE: 2000-08-23
 ; PRIOR APPLICATION NUMBER: US/09/644,600
 ; PRIOR FILING DATE: 1999-10-20
 ; PRIOR APPLICATION NUMBER: 09/027,337
 ; PRIOR FILING DATE: 1998-02-20
 ; NUMBER OF SEQ ID NOS: 98
 ; SEQ ID NO 18
 ; LENGTH: 3147
 ; TYPE: RNA
 ; ORGANISM: Artificial sequence
 ; FEATURE:
 ; OTHER INFORMATION: Antisense of TADG-15
 ; US-09-644-600-18

Query Match 60.6%; Score 1883.2; DB 4; Length 3147;
 Best Local Similarity 81.2%; Pred. No. 0;
 Matches 2223; Conservative 0; Mismatches 508; Indels 5; Gaps 3;
 QY 45 GATCGGACCGCCAAAACCATGGGTAGCAATCGGGGCGCGCAAGCCCGAGGGGCTCTCAG 104
 Db 3143 GAGCGCTCTGGGGTACCATGGGGAGGATCGGGCCGCAAGGGCGAGGGGCGCCGAAG 3084
 QY 105 GATTCGGCGCGGAGCTCAAGTACAACTCCCGCTAGAGAACTGAATGCTTTGAGGAG 164
 Db 3083 GACTTCGGCGCGGAGCTCAAGTACAACTCCCGCACAGAAAGTGAATGCTTTGGAGAA 3024
 QY 165 GGTGTGAGTTCCTGCTCGCAATCCCAAGAAAGTGAGAGCGAGCCCGCAGCGC 224
 Db 3023 GCGTGTGAGTTCCTGCGAGTCAACAGCTCAAGAGGTGAAAGATGATGCGCGGCGC 2964
 QY 225 TGGGTGTGCTGTGGCAGTGTCTTCTGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 284
 Db 2963 TGGGTGTGCTGTGGCAGCGGTGTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 2904
 QY 285 CTGGTGTGGCACTTCCATTCATTCGATGCGGGTTCAAAAGTCTTCATGCGCATCTG 344
 Db 2903 CTGGTGTGGCACTTTCAGTACCGGGACGTGCTGTCTCAGAAAGTCTTCAATGGCTCATG 2844
 QY 345 AGGATCACAATCAGATCTTTCTTGGATCGTATGAGAACTCCACCTCCACAGAGTTTATC 404
 Db 2843 AGATCACAATGAGAAATTTTGGATGCTTACGAACTCCAACTCCACTGATGTTGTA 2784
 QY 405 AGCTTGGCCAGCGAGTGAAGAGGCGCTGAAGCTGCTGTACAAATGAAGTCCCTGCTG 464
 Db 2783 AGCTTGGCCAGCAAGGTGAAGGACGCGCTGAAGCTGCTGTACAGCGAGTCCCATCTCTG 2724
 QY 465 GTTCCCTTACCAAGAGTGGCTGTAACTGCTTCACTGAGGCGAGTGTCACTGCTAC 524
 Db 2723 GGGCCCTTACCAAGAGTGGCTGTGAGCGCTTTCAGCGAGGCGAGCTCATCGCTAC 2664
 QY 525 TACTGTGACAGTTTACAGCATCCCGCCACACCTGCGCAGAGAGTGTGATCGCGCATGGCT 584

Db 2663 TACTGGTCTGAGTTACAGATCCCGCAGACCTGCTGGAGGAGCGCGTCAATGGCC 2604
Qy 585 GTGGAGCGAGTTGTAACTTGGCCACCCGAGCAGCGGCACTGAATCTTCTGCTGCTAACA 644
Db 2603 GAGGAGCGCGTAGTCACTGCTGCCCGCGCGCGCGCTCCCTGAAGTCTTGTGGTCAACC 2544
Qy 645 TCTGTGGTGGCTTCCCATATGACCCAGAAATGCTGCAGAGGACTCAGGACAAACAGCTGC 704
Db 2543 TCAGTGTGGCTTCCCAACGGACTCCAAACAGTACAGAGGACCCAGGACAAACAGCTGC 2484
Qy 705 AGTTTGGCTCGATGCGCATGCTGAGAGTGAACAGCTTCACTACCCCTGGCTTCCCC 764
Db 2483 AGCTTTGGCTCGACGCGCGGCTGTGGAGCTGATGGGCTTCAACAGCGCGGCTTCCCT 2424
Qy 765 AACAGTCCCTACCCGCGCATGCCGCTGCCAGTGGTCTCGGGGGGAGCGGCACTCT 824
Db 2423 GACAGCCCCCTACCCCGCTCATGCGCGCTGCCAGTGGGCGCTTGGGGGGGAGCGGCACTCA 2364
Qy 825 GTGCTGAGCCCTCACCTTCCGAAGCTTGTATGTCTGCTCCCTGTGATGAGCATGGCAGTGAC 884
Db 2363 GTGCTGAGCCCTCACCTTCCGAGCTTGTGACTTGGCTCTGCGACGAGCGCGGAGCGAC 2304
Qy 885 CTGGTACCGTGTATGATGAGCTGAGCCCCCATGGAACCCCAAGCTGTGGTGGCTGTGT 944
Db 2303 CTGGTGAACGCTGTACAAACCCCTGAGCCCCCATGAGCCCCCAAGCTGTGTGCTGTGT 2244
Qy 945 GGCACCTTCTCACCTCCTCAACCTGACTTCTCTCTCCCGAGACGCTTCTTCTTGTTC 1004
Db 2243 GGCACCTTACCTCCTCCTCAACCTGACTTCTCACTCTCCCGAGACGCTTCTGCTCATC 2184
Qy 1005 ACGCTGATAACCAATACTGACCGCGGACATCTTGGCTTGGAGGCACTTCTTCCAGCTG 1064
Db 2183 ACCTGATAACCAACACTGAGCGCGGACATCCCGCTTGGAGGCCACTTCTTCCAGCTG 2124
Qy 1065 CCCAGATGAGCAGCTGTGGGGCTTTTGGAGTGAACACCAAGGGAATTTAGCAGCCCC 1124
Db 2123 CCTAGGATGAGCAGCTGTGGAGGCGCTTACGTAAGCCCCAGGGAATTTCAACAGCCCC 2064
Qy 1125 TACTATCCAGCCCACTACCGCCCAACATCAACTGCACATGGAATCAAGGTGSCCAAC 1184
Db 2063 TACTACCGGCACTACCCACCACTTGAATGACATGACATGCAATGGAATGAGTGGCCCAAC 2004
Qy 1185 AACCGGAACGTGAAGTGGCTTCAAACTCTTCTATCTGGTGGACCCCAACGATACCAAGT 1244
Db 2003 AACCAAGATGTGAAGTGTGAGCTTCAAACTCTTCTATCTGGAGCCCGCGCTGGG 1944
Qy 1245 GCTCTCGACCAAGGATATGTGGAGATCAACGGGAGAGTACTGCGGTGAGAGTCC 1304
Db 1943 GGCACCTGCCCCCAAGGACTACGTGGAGATCAATGGGGAGAAATACTGCGGAGAGAGTCC 1884
Qy 1305 CAGTTTGTGTGAGCAGCAACAGCAAGATTAACGTTCACCTTCCATTTCTGATCACTCG 1364
Db 1883 CAGTTCTGTCTACCAAGCAACAGCAAGATTAACGTTCAGTTCTTCTTCACTAGATCAGTCC 1824
Qy 1365 TACACGAGCAACCGGTTCTTAGTGAATACCTCTTCTACGACTCCCAACGACCGGTGCCCA 1424
Db 1823 TACACCGACACCGGCTTCTTAGTGAATACCTCTTCTACGACTCCCAACGACCGATGCCG 1764
Qy 1425 GGGATGTTATGTGCAAGACTGAGCGTGTGATCCGAAGGAACTGCGCTGGAGCGGTGG 1484
Db 1763 GGGCAGTTTACGTGCCGACCGGGCGGTGTATCCGGAAGGAGCTGCGCTGTGATGGCTGG 1704
Qy 1485 GCAGACTGCCCGGATTAATAGTATGAGCGTTTACTTGGCGGATGCAATGCCACCCAGTTC 1544
Db 1703 GCGACTGACCCAGCAGGAGTGAATGATGATGATGATGATGATGATGATGATGATGATGAT 1644
Qy 1545 ACGTGCAAAACACAGTTCTGCAAGCCCTCTTCTGGGTCTGTGACAGTGTCAACAGCTGT 1604
Db 1643 ACGTGCAAAACAGTTCTGCAAGCCCTCTTCTGGGTCTGTGACAGTGTGAAACAGCTGT 1584
Qy 1605 GGGGACGGAAGTGAACGAGGAGGCTGACGCTGCTCTGCTGGAGTTCACAGTGTTCAT 1664
Db 1583 GGAACAAACGCGACGAGCGGGGTGAGTTGTCTCCGCCCCAGACCTTTCAGTGTTCAT 1524

Qy 1665 GGAAGTGTCTCCCTCAGAGCCAGAGAGTGAATGGGAGAGCAACTGTGGAGATGGTCT 1724
Db 1523 GGAAGTGTCTCTCGAAAAGCCAGAGTGAATGGGAGAGCAACTGTGGGAGCGGTCC 1464
Qy 1725 GAGAGGCTTCATGTGACAGCGTGAATGCTCTCTTGGACCAAAATATACCTACCGCTGC 1784
Db 1463 GACGAGGCTCTCTGCCCAAGGTGAAGCTGCTCACTTGTACAAAACACACTACCGTGC 1404
Qy 1785 CAAAATGGCTCTGTGTGAGCAAGGGCAACCCCTGAGTGTGATGGGAGAGCGGACTGTAGC 1844
Db 1403 CTCAATGGGCTCTCTGTGAGCAAGGGCAACCCCTGAGTGTGACGGAGAGGAGTGTAGC 1344
Qy 1845 GATGGCTCCCATGAGAAAACCTGCTGTGGCTGCGATCCTTTTACCAACAGGCTGC 1904
Db 1343 GACGGCTCAGATGAGAGAGGACTGCGACTGTGGGCTGGGTCATTCAGAGACAGGCTCGT 1284
Qy 1905 GTGTTGGTGGCAGCAATGCGACGAGGGGAGTGGCCCTGGCAGGTGAGCTCCACGCC 1964
Db 1283 GTTGTGGGGGACGATGCGGATGAGGGGAGTGGCCCTGGCAGGTAAAGCTGCATGCT 1224
Qy 1965 CTGGGCCAGGGCCACTTGTGTGGGGCTCGCTCATCTCTCTGACTGGCTGGTCTTGCA 2024
Db 1223 CTGGGCCAGGGCCACTTGTGGGTCTTCTCATCTCTCCCACTGGCTGGTCTCTGCC 1164
Qy 2025 GCTCATTTGCTTTCAGGATGACAAAATTTCAAGTACTCAGACTACACGATGATGGACGGCC 2084
Db 1163 GCACCTGCTTACATCGATGACAGAGGANTCAGGTACTCAGACCCCAAGAGTGGACGGCC 1104
Qy 2085 TTCTGGGCTGTCTGGACCAAGAGCGCAGTGGCTCTGGGGTGCAGGAGCTGAAGCTC 2144
Db 1103 TTCTGGGCTGTGACGACCAAGAGCGCAGCGCCCTGGGCTGCAGGAGCGCAGGCTC 1044
Qy 2145 AAACGTATCATACCCACCCCTTCTTCAATGATTTCACTTGAATGATGATGATGATGATGAT 2204
Db 1043 AAACGTATCATATCTTCCACCCCTTCTTCAATGATTTCACTTGAATGATGATGATGATGAT 984
Qy 2205 CTGAGCTGAGAGAGTCTGGTGGAGTACAGCACCGTCTGTGGCGCCCATCTCCCTGGCTGAT 2264
Db 983 CTGAGCTGAGAGAAACCGGACAGTACAGCTCATGTTGGCGGCCCATCTGCTTGGCGGAC 924
Qy 2265 GCTACCCATGTCTTCCCTGTCTGCAAGGCGCATCTGGGTCAAGGCTGGGGGCAACAAAA 2324
Db 923 GCTCTCCATGTCTTCTTCTGCGGCAAGGCCATCTGGGTCAAGGCTGGGGGCAACACCAG 864
Qy 2325 GAGGAGGTACCGAGCGCTGATCTGCGAGAAAGGTGAGATCCGTGTCTCATCAACAGAC 2384
Db 863 TATGAGGCACTGGCGGCTGATCTCTGCAAAAGGCTGAGATCCGCGCTCATCAACAGAC 804
Qy 2385 ACCTGTGAGGACCTCATCGCGCAGCAGATCACCCCAAGGATGATGATGATGATGATGATGAT 2444
Db 803 ACCTGTGAGGACCTCATCGCGCAGCAGATCACCGCGGATGATGATGATGATGATGATGAT 744
Qy 2445 AGTGGGGTGTGATCTCTGCGCAGGCTGATCTGTGGGCGCCCTTGTCAAGCGCGGAGAAA 2504
Db 743 AGCGGGCGGTGACTCTCTGCGCAGGCTGATCTGCGGGGACCCCTGTCCAGCGTGGAGCG 684
Qy 2505 GATGGCGATGTTCCAGGCTGGTGTGATGCTGGGTGAAGGCTGCGCTGAGAGGAAAC 2564
Db 683 GATGGCGATGTTCCAGGCGGCTGTGGTGTGATGCTGGGAGACGGCTGCGCTGAGAGGAAAC 624
Qy 2565 AAGCCAGGCGGTGACAAAGGCTCCCTGTAGTTCTGGGACTGGATCAAGAGCAGCACTGGG 2624
Db 623 AAGCCAGGCGGTGACAAAGGCTCCCTGTGTTCTGGGACTGGATCAAGAGGAAACCTGGG 564
Qy 2625 GTATAGCAGCATGGACAGACGCGGACCAAAACACCCACAGGAGTGCCTGCAATGCACA 2684
Db 563 GTATAGGCGCGGG---GCCACCAATGTGTACCTGCGGGGCGCACCATCGTCCACC 507
Qy 2685 CCTGGATACAGGAGGAAACACTGACATTTATGCTGTGGCTCCCCCCCCCAACACA 2744
Db 506 CCAGTGTGACG---CTTGGAGGCTGGAGACTGGACCGCTGACTGACACAGCGGCC---CCAGA 449

QY 2745 ACCGAGACTGTGAACATCCTTTAGGACTCAGAGT 2780
|||
Db 448 ACATACACTGTGAACTCAATCTCCAGGGCTCCAAAT 413

RESULT 4

US-09-027-337-9

; Sequence 9, Application US/09027337B

; Patent No. 5972616

; GENERAL INFORMATION:

; APPLICANT: O'Brien, Timothy J.

; APPLICANT: Tanimoto, Hirotooshi

; TITLE OF INVENTION: TADG-15: An Extracellular Serine Protease Overexpressed in

; TITLE OF INVENTION: Breast and Ovarian Carcinomas

; FILE REFERENCE: D6064

; CURRENT APPLICATION NUMBER: US/09/027,337B

; CURRENT FILING DATE: 1998-02-20

; NUMBER OF SEQ ID NOS: 13

; SEQ ID NO 9

; LENGTH: 2900

; TYPE: DNA

; ORGANISM: Homo sapiens

; FEATURE:

; OTHER INFORMATION: SNC19 mRNA sequence (U20428)

US-09-027-337-9

Query Match 49.3%; Score 1530.8; DB 2; Length 2900;
Best Local Similarity 79.1%; Pred. No. 0;
Matches 2011; Conservative 0; Mismatches 502; Indels 29; Gaps 15;

QY 222 CGCTGGTGTGTGGTGGAGTGTCTTTCAGCTTCCTCTTGTCTTCCCTCATGGCTGGC 281

Db 1 CGCTGGTGTGTGTGGAGCGGTGTGATCGGCTCCTCTTGTCTTGTCTGGGATCGGC 60

QY 282 TTGCTGTGTGGTGGACATTCATATCGAATGTGGGTTCAAAAGTCTTCAATGGCCAT 341

Db 61 TTCTGTGTGGTGGATTTGCAATACCGGAGTGTGGTGTCCAGAAAGTCTTCAATGGCTAC 120

QY 342 CTGAGGATCAAAATGAGATCTTTCTGATGCGTATGAGAACTCCACCTCCACAGAGTTT 401

Db 121 ATGAGGATCAAAATGAGAAATTTTGTGATGCTCTACGAACTCCAACTCCACTGAGTTT 180

QY 402 ATCAGCTGGCCAGCCAGGTGAAGAGCGCTGAAGCTGTGTGTAATGAATGCCCTGTTC 461

Db 181 GTAAGCTGGCCAGCAAGGTGAAGAGCGCTGAAGCTGTGTGTAATGAATGCCCTGTTC 240

QY 462 CTGGGTCCCTACACAGAGTGTGGTGTACTGCTTCACTGAGGAGTGTGATCGCC 521

Db 241 CTGGGCCCCCTACCAAGAGTGTGGTGTGACGGCTTTCAGCGAGGCGAGGTTCATCGCC 300

QY 522 TACTACTGGTTCAGAGTTTCAGATCCCCCACCACCTGGCAGAAAGTGTGATCGCGCCATG 581

Db 301 TACTACTGGTTCAGATTCAGATCCCCCAGCACCTGGTGTGAGGAGCGGAGCGGTCTATG 360

QY 582 GCTGTGAGCGAGTTGTAACTATGCGACCCCGAGCAGCGGCACTGAAATCTTCGTGCTA 641

Db 361 GC-CAGGAGCGGTGATGATCTGCTCCCGCGGGCGGCTCCCTGAAATCTTCCTTGTGCTC 419

QY 642 ACATCTGTGTGGCTTCCCATTTGACCCAGAAATGCTGAGAGGACTTCAGGACAAACAGC 701

Db 420 ACCTAGTGTGTGGCTTTCACCGAGCTCCAAACAGTACAGAGGACCCAGGACAAACAGC 479

QY 702 TGCAGTTTGGCTTCGATCCCATGTTGTCAGTGTGACAGCTTCCTTCACTTACCCCTGGCTTC 761

Db 480 TGCAGTTTGGCTTCGATG-CCGCGGTGTGAGGTGATGCGCTTCCACACGCC-GGCTTC 537

QY 762 CCCAACAGTCCCTACCCGGCGATGCCCGTGTCCAGAGTGGTCTCTGGGGGGGAGCGCCGAC 821

Db 538 CTTGACAGCCCTTACCCCGCTCATGCCCGCTGCCAGTGG-----GCTCGGGGAGCGGAC 592

QY 822 TCTGTGTGAGCCCTACCTTCCGAGCTTTGATGTGCTTCCCTGTGATGATGAGTGGCAGT 881

Db 593 GCAGTGTGAGCTACTCGAGCTGACTTCGACGCTTGAATGTGGCGCTTCGAGCGCGGCGAGC 652

QY 882 GACCTGGTCCACGCTGTATGATAGCCTGAGCCCATGAAACCCCAACGCTGTGGTGGCGCTG 941
Db 653 GACCTGGTGA-CGTGTACAAACACCTTGAGCCCCATGGAGCCCCACGC---CTGGTGAAGT 708
QY 942 TGTGGCACCTTCTTACCCCTCCTACAACTGACTTTTCTCTCC--TCCAGAAAGTCTTCC 999
Db 709 TGTGGCACCTTACCTCCTCCTTACAACTGACCTTCCACTCCTCCCAACGAACTCTGTC 768
QY 1000 TTGTACAGCTGATAACCAATACTGACCGGGGACATCTTGGCTTTGAGGCCACTTTCTTC 1059
Db 769 TCATCACACTGTAAACCAACACTGAC--CGGGCATCCGGCTTTGAGGCCACCTTCTTCC 826
QY 1060 AGCTGCCCAAGATGACGAGCTGTGGCGGCTTTTGTAGTGCACCCCAAGGAGACATTTAGCA 1119
Db 827 AGCTGCTTAGATGACGAGCTGTGGAGGCGCTTAGTAAAGCCAGGGAGACATTCAACA 886
QY 1120 GCCCTTACTATCCAGGCCACTACCCGCCCAACATCAACTGCACATGAAATPATCAAGTGC 1179
Db 887 GCCCTTACTATCCAGGCCACTACCCCAACCAATGACTGACATGCAATGMAAATTTAGTGC 946
QY 1180 CCAACACCGGAGCTGAAGTGGCTTCAAACTCTTCTATCTGTGGACCCCAAGCTAC 1239
Db 947 CCAACACCGAGCATGTGAAGTGGCTTCAAACTCTTCTTCTACCTGTGGAGCCCGGGTGC 1006
QY 1240 CAGTGGGCTCCTGCACCAAGGACTATGTGGAGATCAACGGGGAGAAAGTACTGCGGTGAGA 1299
Db 1007 CTGCGGGCACTGCCCCNAGGACTAGTGGAGATCAATGGGAGAAATACTGCGGAGAGA 1066
QY 1300 GGTCCAGATTTGTGTGAGCAGCAACAGCAGCAAGATTAAGTCCACTTCAATCTGATC 1359
Db 1067 GGTCCAGATTTGTGTGAGCAGCAACAGCAGCAAGATTAAGTCCACTTCAATCTGATC 1126
QY 1360 ACTGTGTACAGGACACCGGCTTCTAGCTGAGTACCTCTCTACGACTCCCAAGACCCGT 1419
Db 1127 AGTCTTACCCGACACCGGCTTCTTAGCTGAATACCTCTCTAGACTCCAGTGAACCAT 1186
QY 1420 GCCCAGGAGTGTTCATGTGCAAGACTGGAGCGTGCATCCGAAAGAACTGCGCTGCGACG 1479
Db 1187 GCCCGGGCGAGTTTCACTGCGCACCGGGCGGTGTATCCGGAAGGAGCTGCGCTGTGATG 1246
QY 1480 GCTGGGAGACTCCCGGATTAATGATGATGAGGCTTACTGCCGATGCAATGCCACCCACC 1539
Db 1247 GCTGGGG-GACTGCACCGCACAGCGATGAGTCAACTGCACTGCGACCGCGGCCACC 1305
QY 1540 AGTTTCACTGCAAAAACCACTTCTGCAAGCCCTCTTCTGGGTCTGTGACAGTGTCAACG 1599
Db 1306 AGTTTCACTGCAAGAGCAAGTTCTGCAAG---CTCTTCTGGGTCTGCGACAGTGTGACG 1362
QY 1600 ACTGTGGGAGCGGAAGTGAACGAGGAGGCTGCAAGTGTCTCTG-CTGGGAGTTTCAAAGTGT 1658
Db 1363 AGTGGGAGACAACAGCAGCAGCGAGCGGTTGCAATTTGTCGGGACCCAGACCTTTCAGGTGT 1422
QY 1659 TCCAATGGGAAGTGTCTCCCTCAGAGCCAGCAAGTGTAAATGGGAAGCACTGTGGAGAT 1718
Db 1423 TCCAATGGGAAGTGTCTCCCTCAGAGCCAGCAAGTGTAAATGGGAAGCACTGTGGGGAC 1482
QY 1719 GGGTCTGACGAGGCTTCATGTGACAGCGTGAATGTGCTCTTTGACCAAAATATACCTAC 1778
Db 1483 GGGTCTGACGAGGCTTCCTGCCCAAGGTGAAGTGTGCTACTTTGTACCAACACACCTAC 1542
QY 1779 CGTGTCCAAATGGCTCTGTCTGAGCAAGGGCAACCTCGAGTGTGATGGGAAGACGGAC 1838
Db 1543 CGTGTCTCAATGGGCTCTGCTTTGAGCAAGGGCAACCTCGAGTGTGACGGGAAGGAGAC 1602
QY 1839 TGTAGGATGGCTCCGATGAGAAAAACTGTGACTGTGGGCTGCGATCCTTTACCAACAG 1898
Db 1603 TGTAGCAGCGCTCAGATGAGAAAGCACTGCACTGTGGGCTGCGCTCAATTCACGAGACAG 1662
QY 1899 GCTCGGCTGTGGTGGCAGAAATGCGGACGAGGCGAGTGGCCCTGGGAGGAGTGGAGCTC 1958
Db 1663 GCTCGTGTGTTGGGGGACCGGATGGGATGAGGGCGAGTGGCCCTGGCAGGTAAGCTGT 1722

1959 CAGCCCTGGCCAGGGCCATTGTGTGGGGCTCGCTCATCTCTCTGACTGGCTGTC 2018
1723 CATGCTCTGGCCAGGGCCCATCTGGGTGCTTCCCTCATCTCTCCAACTGGCTGTC 1782
2019 TCTGCACTCATCTGCTTTCAGGATGACAAAATTTCAAGTACTCAGACTACAGATGG 2078
1783 TCTGCCCACTACTACTCATCTGATGACAGAGATTCAGGTACTCAGACCCACG--CAGG 1840
2079 ACGCCCTTCTGGTCTGTGGACGACAGAGCAAGCGAGTCTCTGGGTGTCAGAGCTG 2138
1841 ACGCCCTTCTGGCTGTGGACGACAGAGCAAGCGAGTCTCTGGGTGTCAGAGCTG 1898
2139 AAGCTCAAAAGTATCAGCAAGCCCTTCTCTCAATGATTTCACTTCCGATGATC 2198
1899 AGGCTCAAGCGCATCATCTCCACCCCTTCTTCAATGATTTCACTTCCGATGATC 1958
2199 GCTTGTCTGGAGCTGGAGAGTGGTGGAGTACAGCACCGTCTGTCGCCCCATCTGCTG 2258
1959 GCGCTGTGGAGCTGGAGAACCGGACAGGTACAGTCCATGGTGGCGCCATCTGCTG 2018
2259 CCGTGTACCTATGCTTCTCTGCTGGGAGGCAAGGCAAGTCTGGGTGTCAGAGTGGG 2318
2019 CCGAGCGCTCCATGCTTCTCTGCTGGGCAAGGCAAGTCTGGGTGTCAGAGTGGG 2078
2319 ACAAAGAGGGAGTACCGAGCGCTGATCTCTGAGAGGCTGAGATCCGCTGTCATCAAC 2378
2079 ACCAGATGAGGAGCTGCGCGCTGATCTCTGCAAAAGGCTGAGATCCGCTGTCATCAAC 2138
2379 CAGACCACTGTGAGGAGCTCATCGCGAGAGATCAACCCCAAGATGATGTGTGGGT 2438
2139 CAGACCACTGTGAGGAGCTCATCGCGAGAGATCAACCCCAAGATGATGTGTGGGT 2198
2439 TTCTCTAGTGGGGTGTGGACTCTGCGAGAGTACTCTGTGTGGTGGTGGTGGTGGT 2498
2199 TTCTCTAGCGCGGTGGACTCTGCGAGGCTGATTCGCGGGGACCCCTGTCAGCGTG 2258
2499 GAGAAAGATGGCGAATGTTCCAGCTGTGTGGTGGTGGTGGTGGTGGTGGTGGTGG 2558
2259 GAGCGGATGGCGGATCTTCCAGCGGCTGTGTGGTGGTGGTGGTGGTGGTGGTGGT 2317
2559 AGGAACAGCGCGGTGTACAAAGGCTCTCTGTGTGGTGGTGGTGGTGGTGGTGGTGG 2618
2318 AGGAACAGCGCGGTGTACAAAGGCTCTCTGTGTGGTGGTGGTGGTGGTGGTGGTGG 2377
2619 ACTGGGTATAGCAGTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2678
2378 ACTGGGTATAGGAGCGCGG---GCCACCAAAATGTGTACACCTGCGGGGCAACCCATCG 2434
2679 TGCACACCTGGATCAG 2738
2435 TCCACCCAGTGTGACCGCTGAGGCTGGAGACTCGCGCACCGTGACCTGCACCGAGCG 2494
2739 AACACACCGAGTGTGAAT 2760
2495 CCAGAACATACACTGTGAAT 2516

RESULT 5

US-09-644-600-9

Sequence 9, Application US/09644600

Patent No. 6451500

GENERAL INFORMATION:

APPLICANT: O'Brien, Timothy J.

APPLICANT: Tanimoto, Hiroshi

TITLE OF INVENTION: TADG-15: An Extracellular Serine Protease

FILE REFERENCE: D6064CIP/D

CURRENT APPLICATION NUMBER: US/09/644,600

CURRENT FILING DATE: 2000-08-23

PRIOR APPLICATION NUMBER: 09/421,213

PRIOR FILING DATE: 1999-10-20

PRIOR APPLICATION NUMBER: 09/027,337

PRIOR FILING DATE: 1998-02-20

NUMBER OF SEQ ID NOS: 98
SEQ ID NO 9
LENGTH: 2900
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: SNC-19; GeneBank Accession No. 6451500 #U20428
US-09-644-600-9

Query Match 49.3%; Score 1530.8; DB 4; Length 2900;
Best Local Similarity 79.1%; Pred. No. 0;
Matches 2011; Conservative 0; Mismatches 502; Indels 29; Gaps 15;

222 CGTGGGTGTGTGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 281
Db 1 CGTGGGTGTGTGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 60
282 TTGCTGTGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 341
Db 61 TTGCTGTGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 120
342 CTGAGGATCAAAATGAGATCTTTCTGGATGCGTATGAGAACTCCACCTCCACAGAGTTT 401
Db 121 ATGAGGATCAAAATGAGAAATTTTGTGGATGCGCTTACGAGAACTCCAACTCCACTGAGTTT 180
402 ATCAGCTGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 461
Db 181 GTAAAGCTGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 240
462 CTGGTGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 521
Db 241 CTGGTGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 300
522 TACTACTGTGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 581
Db 301 TACTACTGTGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 360
582 GCTGTGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 641
Db 361 GC-CAGAGCGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 419
642 ACATCTGTGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 701
Db 420 ACCTCAGTGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 479
702 TGCAATTTTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 761
Db 480 TGCAGCTTTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 537
762 CCCAACAGTTCCTTACCGGGCGCATGCGCGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 821
Db 538 CCTGACAGCCCTTACCGGGCGCATGCGCGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 592
822 TCTGTGTGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 881
Db 593 GCAGTGTGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 652
882 GACCTGTGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 941
Db 653 GACCTGTGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 708
942 TGTGGTGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 999
Db 709 TGTGGTGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 768
1000 TTGTGCTGTGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 1059
Db 769 TCATCAGTGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 826
1060 AGCTGCTGTGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 1119
Db 827 AGCTGCTGTGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 886

1120	QY	GC	CCCTACTAT	CCAGGCCACTACCGGCCCAACATCAACTGTCACATGGAATATCAAGTGC	1179
887	Db	GC	CCCTACTAT	CCAGGCCACTACCCACCACCAATTGACTGCACATGGAATAATTCAGGTGC	946
1180	QY	CC	AACACCGGAACGTGAAGTGGCCTTCAAACTCTTTATCTGGTGACCCCAACGTCAC	1239	
947	Db	CC	AACACCGCATGTGAAGTGGCCTTCAAAATCTTCTACCTGTGAGCCCGCGGTGC	1006	
1240	QY	CAGT	GGGCTCCTGCACCAAGGACTATGTGGAGATCAACGGGGAGAACTACTGCGGTGAGA	1299	
1007	Db	CT	CGGGGACCTGCCCCAAGGACTAGCTGGAGATCAATGGGGAGAAATACTGCGGAGAGA	1066	
1300	QY	GGT	CCCAAGTTTGTGTGACGACCAACAGCAGCAAGATTACAGTCACTTCATTTCTGATC	1359	
1067	Db	GGT	CCCAAGTTTGTGTGACGACCAACAGCAGCAAGATTACAGTCACTTCATTTCTGATC	1126	
1360	QY	ACT	GTGTACAGGACACGGGTTCTTAGCTAGTACTCTCTTACGACTTCCAAAGACCCGT	1419	
1127	Db	AGT	CTTACACCGGACACCGGTTCTTAGCTAGTAAATACCTCTTCTTACGACTTCCAGTGA	1186	
1420	QY	GCC	CAGGATGTTTCATGTGCAAGACTTGACGGTGCATCCGAAGGAACCTGGCTGCGACG	1479	
1187	Db	GCC	CGGGCAGTTTCACTGCGCGACCGGGGGTGTATCCGAAGGAGGTGGCTGTGATG	1246	
1480	QY	GCT	GGGCAGACTGCCCGGATTAATGATGATGAGGTTACTGCGCGATGCAATGCCACCCACC	1539	
1247	Db	GCT	GGGC-GACTGCACCGACCAACGCGATGAGCTCAACTGCAATTCGACGCGCGGCCACC	1305	
1540	QY	AGT	TACGTCGCAAAACACAGTTCTGCAAGCCCTCTTCTGGGTCTGTGACAGTGTCAACG	1599	
1306	Db	AGT	TACGTCGTCAGAGCAAGTTCTGCAAG--CTCTTCTGGGTCTGCGACAGTGTGAACG	1362	
1600	QY	ACT	GTGGGACCGAAGTGCAGGAGGGCTGACGCTGTCTGTG--CTGGGAGTTTCAAGTGT	1658	
1363	Db	AGT	CGGAGACCAACAGCAGCAGGAGGTTGCAATTTGTCGGACCCAGACCTTCAAGTGT	1422	
1659	QY	TCC	AATGGGAAGTGTCTCCCTCAGAGCCAGCAAGTGTAAATGGGNAAGCAACACTGTGGAT	1718	
1423	Db	TCC	AATGGGAAGTGTCTCTGAAAGCCACAGTGCATATGGGAAGACGACTGTGGGAC	1482	
1719	QY	GGT	CTCAGAGGCTTCATGTGCAGCGTGAATGTGCTCTCTTGCAACCAATATACCTAC	1778	
1483	Db	GGT	CTCAGAGGCTTCATGCCCCAAGTGAACGTGTCTATTTGATACCAACACACCTAC	1542	
1779	QY	CGT	GCCAAAAATGGCCTGTGTGACCAAGGCAACCCCTGAGTGTGATGGGAAGACGGAC	1838	
1543	Db	CGT	GCCCTCAATGGCTCTGTTGAGCAAGGCCAACCCCTGAGTGTGACGGGAAGGAGGAC	1602	
1839	QY	TGT	AGCGATGGCTCGATGAGAAAACTGTGACTGTGGCTGGCATCTTTTACCAACAG	1898	
1603	Db	TGT	AGCGACGGCTCAGATGGAAGGACTGTGGACTGTGGGCTGGGCTCATTTACGAGACAG	1662	
1899	QY	GCT	CGCGTGTGTGGTGGCAACAAATCCGACGAGGGCGAGTGGCCCTGGCAGGTGAGCCTC	1958	
1663	Db	GCT	CGTGTGTGTGGGGCACCGATGCGGATGAGGGCGAGTGGCCCTGGCAGGTAAAGCCTG	1722	
1959	QY	CAC	GCCTGGGCCAGGGCCACTGTGTGGGCGCTGCTCATCTCTCTGACTGGCTGGTGC	2018	
1723	Db	CAT	GCTCTGGGCCAGGGCCACATCTGCGGTGCTTCCCTCATCTCTCCCAACTGGCTGGTC	1782	
2019	QY	TCT	GACGCTCATTTGTTTCAGGATGACAAAAATTTCAAGTACTCAGACTACAGCATGTGG	2078	
1783	Db	TCT	GCGCACACTGTCTACATCATGATACAGAGGATTCAGTGTACTCAGACCCACG--CAGG	1840	
2079	QY	ACG	GCCTTCTGGGTCTGCTGTGACCAAGCAAGCGAGTGGCTCTGGGTGTCAGAGGCTG	2138	
1841	Db	ACG	GCCTTCTGGGTCTGACACACAGAGCCAGCGCA--GGCCCTGGGGTTCAGAGCGC	1898	
2139	QY	AAG	CTCAAAACGTATCATCAACCAACCTTCTCTTCAATGATTTCACTTGCATTTGACATC	2198	
1899	Db	AGG	CTCAAGGCATCATCTCCCCACCCCTTCTTCAATGACTTCACTTTCGATATGACATC	1958	
2199	QY	GC	CTTGTGAGCTGGAGAAGTGGTGGAGTACAGCACGCTGTCGGCCCATCTGCTGTG	2258	

Db	1959	GGCGTCTGGAGCTGGAGAAACGGGACAGGTACAGCTCCATGGTGGCGCCCATCTGSCCTG	2018
QY	2259	CTGTGATGTACCCATGTCTTCCTCTGGCAAGGCCATCTGGGTACAGGCTGGGGGCAC	2318
Db	2019	CCGAGCGCTGTCATGTCTTCCTCTCCGGCAGGCCATCTGGGTACCGGGCTGGGGACAC	2078
QY	2319	ACAAAAGGAGGTACCGGAGCGGTGATCTCTGCAGAAAGGTGAGATCCGTGTCAATCAAC	2378
Db	2079	ACCCAGTATGGAGGCACCTGGCGCGTGTATCTCTGAAAGGTGAGATCCGCGTCAATCAAC	2138
QY	2379	CAGACCACCTGTGAGGACCTCATATGCCGACAGAGATCACCCCACGAATGATGTGTGGGT	2438
Db	2139	CAGACCACCTGTGAGGACCTCTCTGCCGACAGAGATCAGCGCGCATGATGTGCGTGGGC	2198
QY	2439	TTCTCTCAGTGGGGGTGTGACTCTCTGCCAGGGTGAATCTGTGTGGGCCCTTGTCAAGCGCG	2498
Db	2199	TTCTCTCAGCGCGCGGTGGACTCTCTGCCAGGGTGAATCTCGGGGGGACCCCTGTCCAGCGTG	2258
QY	2499	GAGAAAGATGGGCGAATGTTCCAGGCTGTGTGTGAGCTGGGTGAAGGCTCGCGCTCAG	2558
Db	2259	GAGCGCGATGGGCGGATCTTCCAGGCGGGTGTGTGAGCTGGGAG-ACGCTGGCGTCAAG	2317
QY	2559	AGGAACAAGCCAGGCGGTGTACACAAGGCTCCCTGTAGTTTCGGGACTGGGATCAAGAGCAC	2618
Db	2318	AGGAACAAGCCAGGCGGTGTACACAAGGCTCCCTCTGTTTCGGGAATGGATCAAGAGAAC	2377
QY	2619	ACTGGGGTATAGCAGCATGGACAGACAGCGGACCAACAACCCACAGGATGCCGACA	2678
Db	2378	ACTGGGGTATAGGCGCGGG--GCCACCAATGTGTACACCTGGGGGCCACCCATCG	2434
QY	2679	TGCACACTGGATACAGGAGAGAACTACGACACATTTATGTGTGGCTCTCCCGCCCC	2738
Db	2435	TCCACCCAGTGTGCAGCGCTGCAAGGTGGAGACTCGCGCACCGTGAACCTGCACAGCGC	2494
QY	2739	AACACACCCAGACTGTGAAT	2760
Db	2495	CCGAGAACATACCTGTGAAT	2516
RESULT 6			
US-09-280-116-10			
; Sequence 10, Application 'US/09280116A			
; Patent No. 6331427			
; GENERAL INFORMATION:			
; APPLICANT: Robison, Keith E.			
; TITLE OF INVENTION: Nucleic Acid Molecules Encoding Human Protease Homologs			
; FILE REFERENCE: 5800-24, 035800/176965			
; CURRENT APPLICATION NUMBER: US/09/280,116A			
; CURRENT FILING DATE: 1999-03-26			
; NUMBER OF SEQ ID NOS: 268			
; SOFTWARE: PatentIn Ver. 2.0			
; SEQ ID NO 10			
; LENGTH: 1553			
; TYPE: DNA			
; ORGANISM: Homo sapiens			
; FEATURE:			
; OTHER INFORMATION: trypsin-like serine proteases			
; FEATURE:			
; NAME/KEY: misc_feature			
; LOCATION: (1)..(1553)			
; OTHER INFORMATION: n = a, t, c, or g			
US-09-280-116-10			

RESULTS. 6

```

RES001 6
US-09-280-116-10
; Sequence 10, Application 'US/09280116A
; Patent No. 6331427
; GENERAL INFORMATION:
; APPLICANT: Robison, Keith E.
; TITLE OF INVENTION: Nucleic Acid Molecules Encoding Human Protease Homologs
; FILE REFERENCE: 5800-24, 035800/176965
; CURRENT APPLICATION NUMBER: US/09/280,116A
; CURRENT FILING DATE: 1999-03-26
; NUMBER OF SEQ ID NOS: 268
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 10
; LENGTH: 1553
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: trypsin-like serine proteases
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)..(1553)
; OTHER INFORMATION: n = a, t, c, or g
; US-09-280-116-10

Query March 22.6%; Score 701.2; DB 4; Length 1553;
Best Local Similarity 71.1%; Pred. No. 1-2e-177;
Matches 1092; Conservative 0; Mismatches 351; Indels 92; Gaps 9;

QY 1659 TCCAAATGGGAAGTGTCTCCCTCAGAGCCAGAGTGTAATGGGAAGGACAACACTGTGGAGAT 1718
DB 1 TCCGATGNGAAGTGCCINTTCGAAGAAGCCAGAGTGCAATGGGAAGGACGACTGTGGGGGC 60

QY 1719 GGGTCTGACGGAGGCTTCATGTGACAGCGTGAAATGTCCTCTCTTTCGCCAACAATATACCTAC 1778

```


Db	1136	CTCGTTCCTCAGCCTCCAAAGTGGAGCTGGGAGGTAGAAAGGGAGACACTGGTGGTTC	1195
Qy	2829	TAGCGGCCAGCCTGGGGCAAGGGTTGATGCGAGCTTCCCCCTCTAGCCCTAGACTG	2888
Db	1196	TACTGACCCAACTGGGGGCAAGGTTTGAAGACACAGCCTCCCGCGCAGCCCAAGCTG	1255
Qy	2889	GGTGAAGATGATGCTGTCCCGAGAGCTGCTTC-----CAACTGTCATTGAG	2935
Db	1256	GGCGAGGCGGTTTGTGCATATCTCCCTCCCTGTCTTAAGGAGCAGCGGAAACGGAG	1315
Qy	2936	CTCCGGGAGCCCTA-----TGGAGGAGGGGCTCAGGG	2969
Db	1316	CTTGGGGCTCTCTAGTGAAGTGGTGGGGCTGCCGGATCTGGGCTGTGGGGCTTTGGG	1375
Qy	2970	TCACTCTTTTCAGAGCGCCAGCCCTAGGAAACCCAGAAAGAGTGGTACCTAAGGCTG	3029
Db	1376	CCAGGCTCTTGAGAGGCCCAGGCTCGGAGGACCTTGGAAAAACAGACGGGCTCTGAGACTG	1435
Qy	3030	AAATGTTTGTGCTGTGCCAGGGGTGG-----GTATTGTGAGAGTAAA	3071
Db	1436	AAATTGTTTACCAGCTCCCGAGGTGCACITTCAGTGTGTGTAATTGTGTAATGAGTAAA	1495
Qy	3072	ACATTTTATTTCTTTTAAAAAATAAAAAAAAAA	3106
Db	1496	ACATTTTATTTCTTTTAAAAAATAAAAAAAAAA	1530

RESULT 7

Query Match	9.9%	Score 306.8	DB 4	Length 434	
Best Local Similarity	82.1%	Pred. No. 2e-72			
Matches 353	Conservative	0	Mismatches 77	Indels 0	Gaps 0
Qy	1083	GGCGGCTTTTGTAGTGACACCCAGGGAATTTAGCAGCCCTTACTATCCAGGCCACTAC	1142		
Db	1	GGAGGCCGCTTACGTAAGCCAGCCAGGGGACATTTCAACAGCCCTTACTACCAGGCCACTAC	60		
Qy	1143	CCGCCCAACATCAACTGCGCATGGAAATATCAAGTGCCCAACACCGGAACGTGAAGGTG	1202		
Db	61	CCACCCAAACATTGACTGCGCATGGAAATTTAGAGTGCCCAACAAACAGCAGTGTGAAGGTG	120		
Qy	1203	CGCTTCAAACTCTTCTATCTGTGTGACCCCAACGTTACCACTGGGCTCCTGACCCAAAGGAC	1262		
Db	121	CGCTTCAAATTCCTTACTCTGTGTGAGCCCGGGTGCCCTGCGGCGACCTTGCCCAAGGAC	180		
Qy	1263	TATGTGAGATCAACCGGGGAGAAGTACTCGGTGAGAGGTCCAGTTTGTGTGAGCAGC	1322		
Db	181	TACGTGGAGATCAATGGGAGAAATACTGCGAGAGAGGTCCAGTTCTGTCTGTCACGAC	240		

Query Match

Best Local Similarity 82.1%; Pred. No. 2e-72;

Matches 353; Conservative 0; Mismatches 77; Indels 0; Gaps 0;

Qy 1083 GCGGCTTTTGAGTGACACCAAGGACATTTAGCAGCCCTACTATCCAGGCCACTAC 1142

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
84

DB I GGAGGCCGCTTACGTAAAGCCAGGGGACATTCACACAGCCCCCTACTACCCAGGCCACTAC 60

1143 CCGCCCAACATCAACTGCACTGGAATATCAAGGTGCCCAACAACCGGAACGTGAAGGTG 1202

[illegible]

Db 61 CCACCAACATTGACTGCACATGGAACATTGAGGTGCCAACACAGCATGTGAAGTG 120

QY 1203 CGCTTCAAACCTCTTCTATCTGGTGGACCCCAACGTACCTAGTGGCTCTGCACCAAGGAC 1262

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100

Db 121 CGCTTCAAATTCTTCTACCTGCTGGAGCCCGGGCGTGCCTGCGGGCACCTGCCCAAGGAC 180

1263 TATGTGGAGATCAACGGGGAGAAGTACTGGCGGTGAGAGGTCCTCCAGGTTCTGGTGAGCAGC 1322

QY 1203 TATGTGGAGATCAACGGGGAGAGATACGGGGTGGAGAGGTCACAGTTTGTGGTGAGCAGC 1522

Db 181 TACGTGGAGATCAATGGGGAGAAATACTGGGAGAGAGGTCCAGTTCGTCTCACCAGC 240

QY 1323 AACAGCAGCAAGATTACAGTCCACTTCCATTCTGATCACTGGTACACGACACCGGGTTC 1382
DB 241 AACAGCAACAAGATCACAGTTTCGCTTCCACTCAGATCAGTCTACACGACACCGGCTTC 300
QY 1383 CTAGCTGAGTACCTCTCTTACGACTCCAAAGACCGGTGCCAGGAGTTTCATGTGCAAG 1442
DB 301 TTAGCTGAATACCTCTCTTACGACTCCAGTACCCATGCCGGGGCAGTTTCACGTGCCGC 360
QY 1443 ACTGACCGGTGCATCCGAAAGAACTGCGCTGGACGCGCTGGGAGACTGCCCGGATTAT 1502
DB 361 ACGGGCGGTGTATCCGGAAGAGGTGCGCTGTGATGGCTGGCGGACTGCACCGACAC 420
QY 1503 AGTGATGAGC 1512
DB 421 AGCGATGAGC 430

RESULT 8

US-09-736-457-1480

; Sequence 1480, Application US/09736457

; Patent No. 6509448

; GENERAL INFORMATION:

; APPLICANT: Wang, Tongtong

; APPLICANT: Bangur, Chaitanya S.

; APPLICANT: Lodes, Michael A.

; APPLICANT: Fanger, Gary

; APPLICANT: Vedvick, Tom

; APPLICANT: Carter, Darriek

; APPLICANT: Retter, Marc

; APPLICANT: Mannion, Jane

; APPLICANT: Fan, Liqun

; APPLICANT: Wang, Aijun

; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND

; FILE OF INVENTION: DIAGNOSIS OF LUNG CANCER

; FILE REFERENCE: 210121.478C15

; CURRENT APPLICATION NUMBER: US/09/736,457

; CURRENT FILING DATE: 2000-12-13

; NUMBER OF SEQ ID NOS: 1864

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 1480

; LENGTH: 434

; TYPE: DNA

; ORGANISM: Homo sapien

US-09-736-457-1480

Query Match 9.9%; Score 306.8; DB 4; Length 434;
Best Local Similarity 82.1%; Pred. No. 2e-72;
Matches 353; Conservative 0; Mismatches 77; Indels 0; Gaps 0;

QY 1083 GCGGCTTTTGTAGTGACACCCAGGAGACATTTAGCAGCCCTTACTATCCAGGCCACTAC 1142
DB 1 GAGGGCGCTTACGTAAAGCCCGAGGGAGACATTCACAGCCCTTACTACCCAGGCCACTAC 60
QY 1143 CCGCCCAACTCACTGCACATGAATATCAAGTGGCCCAACACCGGAAAGTGAAGGTG 1202
DB 61 CCACCCAACTGACTGCATATGAACATTTAGGTGGCCCAACACCGAGATGTGAAGGTG 120
QY 1203 CGCTTCAAACTCTTCTATCTGTTGGTGGACCCCAACAGTACAGTGGGCTCTCTGCACCAAGGAC 1262
DB 121 CGCTTCAAACTCTTCTATCTGTTGGTGGACCCCGGCTGCTTGGGGCACTTCCCCCAAGGAC 180
QY 1263 TATGTGAGATCAACGGGAGAGTACTGCGGTGAGAGGTCCCAATTTGTTGGTGAAGC 1322
DB 181 TAOTGGAGATCAATGGGAGAAATACTGCGGAGAGAGTCCCAATTCGTCTGTCACCAAGC 240
QY 1323 AACAGCAGCAAGATTACAGTCCACTTCCATTCTCATCATCTGTCACAGCAGCCGGTTC 1382
DB 241 AACAGCAACAGATCACAGTTTCGCTTCCACTCAGATCAGTCCCTACACCGACACCGGCTTC 300
QY 1383 CTAGCTGAGTACCTCTCTTACGACTCCAAAGCCCGGTGCCAGGAGTGTTCATGTGCAAG 1442
DB 301 TTAGCTGAATACCTCTCTTACGACTCCAGTACCCATGCCGGGGCAGTTTCAGTGGCCG 360

QY 1443 ACTGACGCTGCATCCGAAAGAACTGCGCTGGACGCGTGGGAGACTGCCCGGATTAT 1502
DB 361 ACGGGCGGTGTATCCGGAAGAGGTGCGCTGTGATGGCTGGGCGGACTGCACCGACAC 420
QY 1503 AGTGATGAGC 1512
DB 421 AGCGATGAGC 430

RESULT 9

US-09-280-116-107

; Sequence 107, Application US/09280116A

; Patent No. 6331427

; GENERAL INFORMATION:

; APPLICANT: Robison, Keith E.

; TITLE OF INVENTION: Nucleic Acid Molecules Encoding Human Protease Homologs

; FILE REFERENCE: 5800-24, 035800/176965

; CURRENT APPLICATION NUMBER: US/09/280,116A

; CURRENT FILING DATE: 1999-03-26

; NUMBER OF SEQ ID NOS: 268

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 107

; LENGTH: 796

; TYPE: DNA

; ORGANISM: Homo sapiens

; FEATURE:

; OTHER INFORMATION: trypsin-like serine proteases

; FEATURE:

; NAME/KEY: misc feature

; LOCATION: (1)..(796)

; OTHER INFORMATION: n = a, t, c or g

US-09-280-116-107

Query Match 5.6%; Score 173.2; DB 4; Length 796;

Best Local Similarity 57.7%; Pred. No. 1.4e-36;

Matches 395; Conservative 0; Mismatches 269; Indels 21; Gaps 4;

QY 1744 GCGTAATGTCGTCCTTTCGACCAATATACCTACCGCTGCCAAATGGCTCTGTCTGA 1803
DB 106 GCCAGGAGGGTGCATGTGGACATTCACCTTCAGTGTGAGCCGAGCTGCGTGA 165
QY 1804 GCAAGGGCAACCTGTAGTGTGATGGGAAGACGCACTGTAGCGATGGCTCCGATGAGAAA 1863
DB 166 AGAAGCCCAACCGCAGTGTGATGGGCGGCCGCTGACAGGACGGCTCGATGAGGAGC 225
QY 1864 ACTGTGACTGTGGGCTCGCATCTTTCACCAACAGGCTCGCTGTGGTGGCAGCAATG 1923
DB 226 ACTGTGACTGTGGGCTCGCATCTTTCACCAACAGGCTCGCTGTGGTGGGAGCTGTGT 279
QY 1924 CGGACGAGGGCGAGTGGCCCTTGGCAGGTGAGCCTCCACGCGCTGGGCGGACCTTGT 1983
DB 280 CCTCCGAGGGTGTAGTGGCCATGCGAGGCGAGCTCCAGGTTTCGGGTC---GACACATCT 336
QY 1984 GTGGGGCTCGCTCATCTCTCCGACTGGCTGTGTCTGTGAGCTCATGTCTTCAGGATG 2043
DB 337 GTGGGGGGGCGCTCATCGCTGACCGCTGGGTGATAACAGCTGCCCACTGTCTTCAGGAG 396
QY 2044 ACAAATAATTCAGATGACTCAGACTACAGATGTGAGCGGCTTCTGGGTCTCTGAC 2103
DB 397 ACAGCA-----TGGCCCTCAGCGTGTGTGGACCGGTGTCTTCTGGGCAAGGTGTGGC 447
QY 2104 AGAGCAAGCGCAGTGGCTCTTGGGGTGCAGAGCTGAAGCTCAAACTATCATCACCCACC 2163
DB 448 AGAACTCGGCTG---GCCTGGAGAGGTGTCTTCAAGGTGAGCCGCTCTCTCTGCACC 504
QY 2164 CTTCTTCAATGATTTACCTTCGACTATGACATCGCCTTGTGGAGCTGGAGAGTCGG 2223
DB 505 CGTACCACCAAGAGGACAGCCATGACTAGCAGTGGCGCTGTGTGAGCTGAGCAACCCCG 564
QY 2224 TGGAGTACAGACCGCTGTGGCGCCCATCTGCTGCTGCTGATGCTACCCATGTCTTCCCTG 2283
DB 565 TGGTGGCTCGGGCGCGGTGCGCCCGCTGTGCTGCGCGCGCTCCCACTTCTTCTGAGC 624

QY 2284 CTGGCAAGGCCATCTGGGTCAAGGCTGGGGGCACACAAAGAGGGAGGTACCGGAGCG 2343
DB 625 CCGGCTGCACTGCTGGATTAGGGCTGGGGCGCTTGGCGAGGGCGGCCCATCAGCA 684
QY 2344 TGATCTCAGAGAGGTGAGATCGGTGTCATCAACACAGACACCTGTGAGGACCTCATGC 2403
DB 685 AGCTCTCGAGAAAGTGTGTCANTTGTATCCCAAGGACCTGTGCAGCGAGGTCTATC 744
QY 2404 CGCAGCAGATCACCCACCAAGATGAT 2428
DB 745 GCTACCAGGTGACGCCACGATGCT 769

RESULT 10
US-09-518-046-1
; Sequence 1, Application US/09518046
; Patent No. 6294663
; GENERAL INFORMATION:
; APPLICANT: O'Brien, Timothy J.
; APPLICANT: Underwood, Lowell J.
; TITLE OF INVENTION: Transmembrane Serine Protease Overexpressed
; FILE REFERENCE: D6192CIP
; CURRENT APPLICATION NUMBER: US/09/518,046
; CURRENT FILING DATE: 2000-03-02
; EARLIER APPLICATION NUMBER: 09/261,416
; EARLIER FILING DATE: 1999-03-03
; NUMBER OF SEQ ID NOS: 153
; SEQ ID NO 1
; LENGTH: 2413
; TYPE: DNA
; ORGANISM: Homo sapiens
; NAME/KEY: CDS
; OTHER INFORMATION: entire cDNA sequence of TADG-12 gene
US-09-518-046-1

Query Match 3.7%; Score 113.6; DB 3; Length 2413;
Best Local Similarity 52.6%; Pred. No. 2.1e-20;
Matches 388; Conservative 0; Mismatches 314; Indels 36; Gaps 5;

QY 1900 CTCGGTGGTGGTGGACGAATGGGACGAGGGGAGTGGCCCTGGCAGTGGAGCTTCC 1959
DB 790 CACGCATGCTGGGTGGAAACATGCTTCTCTCGCAGTGGGCCCTTGGCAGCGACGCTTC 849
QY 1960 ACGCCCTGGGCGAGGGCCACTTGTGTGGGGCTCGCTCATCTCTCTGACTGGTGGTCT 2019
DB 850 AGTTCAGGGCTA---CCACTGTGGGGGGCTCTGTCTATCAGCGCCCTGTGGATCATCA 906
QY 2020 CTGAGCTCATTTGCTTTCAGGATGACAAAAATTTCAAGTACTCAGACTACACGATGTGA 2079
DB 907 CTGCTGCACACTGTGTTTATGACTTGTACTTCTCCCAAGTCAATGACCATCCAGTGGTC 966
QY 2080 CGGCTTCTGGTCTGTGACACAGCAGCAGCGAGTGGCTCTGGGGTGCAGAGCTGA 2139
DB 967 TAGTTTCCCTG-----TTGGCAATTCAGCCCATCCCACTTGTGGAGAAG----- 1013
QY 2140 AGCTCAAAAGTATCATCACCCACCTTCTTCAATGATTTTCACTTCGACTATGATATCG 2199
DB 1014 -----ATTGCTTACACAGCAAGTACAGCCAAAGAGCTGGGCAATGACATCG 1062
QY 2200 CTTGTGAGCTGAGAGTGGTGGAGTACAGACCGTCTGGCCGCCCATCTGCGCTGC 2259
DB 1063 CCTTATGAGCTGGCCGGGCACTCACGTTCAATGAAATGATCCAGCTGTGTGCTGTC 1122
QY 2260 CTGATGCTACCATGCTTTCCTCTGTCAGAGGCCATCTGGGTACAGGCTGGGGGACA 2319
DB 1123 CCAACTCTGAAGAGAACTTCCCGCATGGAAAGTGTGCTGGAACGTCAAGATGGGGGCCA 1182
QY 2320 CAAAAGAGGAGGTACCGGAGCGGTGATCTTGCAGAGAGGGTGGAGATCCGTCTCATCAAC 2379
DB 1183 CAGAGGATGGAGGTGACGCTTCCCTCTCTTGAACCAACCGCGGCGCTTCCCTTGTATTTCCA 1242

QY 2380 AGACCACTGTGA-----GGACCTCATGCCGACGACATCACCCACGAATGATGTGTG 2433
DB 1243 ACAAGATCTGCAACACACAGGAGCGGTACGGTGGGATCATCTCCCCCTCCATGCTCTGG 1302
QY 2434 TGGGTTTCTCAGTGGGGGTGTGGACTCTCTCCAGGGTGAATCTCTGTGGCCCCCTTGTCAA 2493
DB 1303 CGGGCTACTCTGACGGTGGCGTGAACAGCTGCCAGGGGGAAGCGGGGGCCCCCTGGTGT 1362
QY 2494 CGCGGAGAAAGATGGCGGAATTTCCAGGCTGTGTGTGGTGTGAGCTGGGGGTGAAGCTGCG 2553
DB 1363 GTCAGAGAGAGGAGGCTGTGGAAGTT--AGTGGAGAGCCAGAGCTTTGGCATCGGCTGCG 1419
QY 2554 CTCAGAGGAACAAGCCAGGCGGTGTACACAAGGCTCCCTGTAGTTGGGACCTGGATCAAG 2613
DB 1420 CAGAGGTGAACAGCCTGGGTGTACACCCGTGTACCTCCTTCTTGACTGGATCCACG 1479
QY 2614 AGCACACTGGGGGTATAGC 2631
DB 1480 AGCAGATGGAGAGAGACC 1497

RESULT 11
US-09-518-046-3
; Sequence 3, Application US/09518046
; Patent No. 6294663
; GENERAL INFORMATION:
; APPLICANT: O'Brien, Timothy J.
; APPLICANT: Underwood, Lowell J.
; TITLE OF INVENTION: Transmembrane Serine Protease Overexpressed
; FILE REFERENCE: D6192CIP
; CURRENT APPLICATION NUMBER: US/09/518,046
; CURRENT FILING DATE: 2000-03-02
; EARLIER APPLICATION NUMBER: 09/261,416
; EARLIER FILING DATE: 1999-03-03
; NUMBER OF SEQ ID NOS: 153
; SEQ ID NO 3
; LENGTH: 2544
; TYPE: DNA
; ORGANISM: Homo sapiens
; NAME/KEY: CDS
; OTHER INFORMATION: entire cDNA sequence of TADG-12 variant gene
US-09-518-046-3

Query Match 3.5%; Score 109.8; DB 3; Length 2544;
Best Local Similarity 54.6%; Pred. No. 2.3e-19;
Matches 266; Conservative 0; Mismatches 212; Indels 9; Gaps 2;

QY 2151 ATCATCACCCACCCCTTCTTCAATGATTTTCACTTCGACTATGACATCGCTTCTGGAG 2210
DB 1145 ATTGTCTACACAGCAAGTACAGCCAAAGAGGTGGGCAATGACATCGCCCTTATGAAG 1204
QY 2211 CTGAGAAAGTGGTGGAGTACAGCACCGTCTGTGGCCCCCATCTGCTGCTGCTGATGTACC 2270
DB 1205 CTGGCCGGGCACTCACGTTTCAATGAAATGATCCAGCCTGTGTGCTGCCCACTCTGAA 1264
QY 2271 CATGTCTTCTCTGCTGCAAGGCCATCTGGGTACAGGCTGGGGGACACAAAGAGGGA 2330
DB 1265 GAGAACTTCCCGATGGAAAGTGTGCTGGACGTGGGGGGCCACAGAGGATGGA 1324
QY 2331 GGTACCGGAGCGTGTATCTCTGCAAGAGGGTGGAGATCCGTGTATCAACACGACCATCTGT 2390
DB 1325 GGTGACGCTTCCCTGTCTTGAACCCAGCGCGCGTCCCTTTGATTTTCCACAGATCTGC 1384
QY 2391 GA-----GGACCTCATGCGCAGCAGATCACCCCAAGATGATGTGTGGTTCCTC 2444
DB 1385 AACACAGGAGCGTGTACGGTGGCATCATCTCCCCCTCCATGTCTCTGCGGGGTACCTG 1444
QY 2445 AGTGGGGGTGTGACTCTCTCCAGGGTGTACTGTGTGGGCCCTTGTCAAGCGGGAGAAA 2504
DB 1445 ACGGTGGCGTGGACAGCTGCCAGGGGGGACAGCGGGGGCCCCCTGGTGTGTCAAGAGAGG 1504

QY 2505 GATGGGGAATGTTCCAGGCTGGTGTGTGAGCTGGGTGAAGCTGGCTCAGAGGAAC 2564
Db 1505 AGGCTGTGGAAGTT--AGTGGAGCGACACAGCTTTGGCATCGGCTGGCGAGAGTGAAC 1561
QY 2565 AAGCCAGGCTGTACACAGGCTCCCTGTAGTTCGGGACTGGATCAAGAGACACACTGGG 2624
Db 1562 AAGCCTGGGTGTACACCGGTGTACCTCTCTCTGACTGGATCCAGGACGACATGGAG 1621
QY 2625 GTATAGC 2631
Db 1622 AGAGACC 1628

RESULT 12

US-09-342-749-1
Sequence 1, Application US/09342749
Patent No. 6166194
GENERAL INFORMATION:
APPLICANT: Worg, Alexander K.C.
APPLICANT: Tavtigian, Sean V.
APPLICANT: Teoq, David H.-F.
APPLICANT: Myriad Genetics, Inc.
TITLE OF INVENTION: TMRSS2 is a Tumor Suppressor
FILE REFERENCE: 2318-202
CURRENT APPLICATION NUMBER: US/09/342,749
CURRENT FILING DATE: 1998-06-29
EARLIER APPLICATION NUMBER: US 60/091,044
EARLIER FILING DATE: 1998-06-29
NUMBER OF SEQ ID NOS: 33
SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 1
LENGTH: 1479
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (1)..(1476)
FEATURE:
NAME/KEY: conflict
LOCATION: (724)
OTHER INFORMATION: Listed as T in GenBank Accession No. U75329
FEATURE:
NAME/KEY: conflict
LOCATION: (985)
OTHER INFORMATION: Listed as C in GenBank Accession No. 6166194 U75329
FEATURE:
NAME/KEY: conflict
LOCATION: (1347)
OTHER INFORMATION: Listed as C in GenBank Accession No. 6166194 U75329
FEATURE:
NAME/KEY: conflict
LOCATION: (1466)
OTHER INFORMATION: Listed as A in GenBank Accession No. 6166194 U75329
FEATURE:
NAME/KEY: conflict
LOCATION: (1471)
OTHER INFORMATION: Listed as A in GenBank Accession No. 6166194 U75329.
FEATURE:
NAME/KEY: allele
LOCATION: (478)
OTHER INFORMATION: This base can be G or A with G being the more
OTHER INFORMATION: common allele. The codon will change from Val to
OTHER INFORMATION: Met.
FEATURE:
NAME/KEY: allele
LOCATION: (777)
OTHER INFORMATION: This base can be C or T with C being the more
OTHER INFORMATION: common allele. The codon is unaffected with both
OTHER INFORMATION: alleles encoding Gly.
FEATURE:
NAME/KEY: allele
LOCATION: (768)

OTHER INFORMATION: This base can be C or T with C being the more
OTHER INFORMATION: common allele. This is a silent polymorphism.
FEATURE:
NAME/KEY: allele
LOCATION: (834)
OTHER INFORMATION: This base can be C or T with C being the more
OTHER INFORMATION: common allele. This is a silent polymorphism.
FEATURE:
NAME/KEY: allele
LOCATION: (625)
OTHER INFORMATION: This base can be T or A with T being the more
OTHER INFORMATION: common allele. The codon will change from Phe to
OTHER INFORMATION: Ile
US-09-342-749-1

Query Match 3.3%; Score 101.6; DB 3; Length 1479;
Best Local Similarity 52.8%; Pred. No. 2,7e-17;
Matches 396; Conservative 0; Mismatches 324; Indels 30; Gaps 7;
QY 1886 CTTTACCAACAGGCTCGCGTGGTGTGGACAGATCGGACGAGGGCGAGTGGCCCTG 1945
Db 747 CTCAAGCGCCAGAGAGGATCGTGGCGCGAGAGCGCGCTCCCGGGGCGCTGGCCCTG 806
QY 1946 GCAGGTGAGCTCCACGCCCTGGGCGCAGGGCCACTTGTGTGGGGCTCTCGTCTCTCTCC 2005
Db 807 GCAGGTGAGCTCCACGGTCCAGAAC---GTCCAGGTGTGGGAGGCTCCATCATCACCCC 863
QY 2006 TGACTGGCTGTCTGTGCAGCTCAITGTCTTCAAGATGACAAAAATTTCAAGTACTCAGA 2065
Db 864 CGAGTGGATCGTACAGCGCCCACTGC-----GTGAAAAACCTCTTAAACAATCCATG 917
QY 2066 CTACAGATGTGACGCGCTTCTGGTCTGTGGACCAAGCAAGCGAGTGCCTCTGG 2125
Db 918 GCAT-----TGGACGGCATTTCCGGGGATTGTAGA-----CAATCTTCATGTTCTAT 966
QY 2126 GTTGAGGAGCTGAAGCTCAACGTATCATCAACCCACCTTCTCTCAATGATTTCACTT 2185
Db 967 GGAGCGGATACCAAG-TAGAAAAAGTGATTTCTCATCCAAATTAATGACTCCCAAGACAA 1025
QY 2186 CGACTATGATCGCTTCTGTGAGCTGGAGAGTGGTGGAGTACAGACCGTCTGGCG 2245
Db 1026 GAAATGATGATTCGCTGATGAAGCTGCAGAGCCTCTGACTTTTCAACGACCTAGTGAA 1085
QY 2246 CCCCATCTGCTGCTGATCTACCATGCTTCCCTGCTGGCAAGCCCATCTGGGTCAC 2305
Db 1086 ACCAGTGTCTGCCCAACCCAGGCGATGCTGCGCCAGACAGCTCTGCTGGATTTC 1145
QY 2306 AGCTGGGGGCACAAAAAGAGGGAGGTA-----CCGAGCGCTGATCTCGCAGAGGG 2359
Db 1146 CGGGTGGGGGGCCACCGAGGAGAAAGGAAGACCTCAGAAAGTCTGAACGCTGCCAAGGT 1205
QY 2360 TGAGATCCGCTGCTCATCAACAGACCACTGTGAGGACCTCATGCCCGCAGATCACCCC 2419
Db 1206 GCTTCTCATTTGAGACACAGATGCAACAGCAGATATGTTATGACAACTGATACACC 1265
QY 2420 ACGAATGATGTGTGGGTTTCTCAGTGGGGGTGTGGACTCCTGCCAGGGTGACTCTGG 2479
Db 1266 AGCATGATCTGTGCCGGCTTCTTGAGGGGAACTCGATTCTTGGCAGGTCACAGTGG 1325
QY 2480 TGGCCCTTGTCAAGCGCGAGAGAAATGGGCGAATGTTTCCAGGCTGGTGTGGTGAAGTGG 2539
Db 1326 AGGGCTCTGTCTCACTTCGAAGAAACAATATCTG--GTGGCTGATAGGGGATACAGCTG 1382
QY 2540 GGTGAAGCTGTGCTCAGAGGAACAAGCGCGTGTACACAGGCTCCCTGTAGTTCG 2599
Db 1383 GGGTCTGCTGTGCCAAAGCTTACAGACAGAGTGTACGGGAATGTGATGATTTCAC 1442
QY 2600 GGAATGATCAAGAGACACACTGGGGTATA 2629
Db 1443 GGAATGATTTATCGACAAATGAGGGCAGA 1472

RESULT 13

US-09-691-840-1
; Sequence 1, Application US/09691840
; Patent No. 6444419
; GENERAL INFORMATION:
; APPLICANT: Wong, Alexander K.C.
; APPLICANT: Tavtighian, Sean V.
; APPLICANT: Teng, David H.-F.
; APPLICANT: Myriad Genetics, Inc.
; TITLE OF INVENTION: TWPRS2 is a Tumor Suppressor
; FILE REFERENCE: 2318-202
; CURRENT APPLICATION NUMBER: US/09/691,840
; CURRENT FILING DATE: 2000-10-18
; PRIOR APPLICATION NUMBER: US/09/342,749
; PRIOR FILING DATE: 1999-06-29
; PRIOR APPLICATION NUMBER: US 60/091,044
; PRIOR FILING DATE: 1998-06-29
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1

LENGTH: 1479
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (1)..(1476)
NAME/KEY: conflict
LOCATION: (724)
OTHER INFORMATION: Listed as T in GenBank Accession No. U75329
NAME/KEY: conflict
LOCATION: (985)
OTHER INFORMATION: Listed as C in GenBank Accession No. 6444419 U75329
NAME/KEY: conflict
LOCATION: (1347)
OTHER INFORMATION: Listed as C in GenBank Accession No. 6444419 U75329
NAME/KEY: conflict
LOCATION: (1466)
OTHER INFORMATION: Listed as A in GenBank Accession No. 6444419 U75329
NAME/KEY: conflict
LOCATION: (1471)
OTHER INFORMATION: Listed as A in GenBank Accession No. 6444419 U75329.
LOCATION: (478)
OTHER INFORMATION: This base can be G or A with G being the more
OTHER INFORMATION: common allele. The codon will change from Val to
OTHER INFORMATION: Met.
NAME/KEY: allele
LOCATION: (777)
OTHER INFORMATION: This base can be C or T with C being the more
OTHER INFORMATION: common allele. The codon is unaffected with both
OTHER INFORMATION: alleles encoding Gly.
NAME/KEY: allele
LOCATION: (768)
OTHER INFORMATION: This base can be C or T with C being the more
OTHER INFORMATION: common allele. This is a silent polymorphism.
NAME/KEY: allele
LOCATION: (834)
OTHER INFORMATION: This base can be C or T with C being the more
OTHER INFORMATION: common allele. This is a silent polymorphism.
NAME/KEY: allele
LOCATION: (625)
OTHER INFORMATION: This base can be T or A with T being the more
OTHER INFORMATION: common allele. The codon will change from Phe to Ile
US-09-691-840-1

Query Match 3.3%; Score 101.6; DB 4; Length 1479;
Best Local Similarity 52.8%; Pred. No. 2.7e-17;
Matches 396; Conservative 0; Mismatches 324; Indels 30; Gaps 7;
QY 1886 CTTTACCAACAGGCTCGGTGGTTGGTGGACGATCGGACGAGGCGGAGTGGCCCTG 1945
Db 747 CTCAGCGCGCAGACGAGATCGTGGGGCGGAGAGCGCTCCCGGGGCGCTGGCCCTG 806
QY 1946 GCAGGTGAGCCTCCACGCGCTGGGCGACCTTGTGTGGGGCCTCGCTCATCTCTCC 2005

Db 807 GCAGGTGAGCCTCGACGTCCAGAAC---GTCCAGCTGTGGGAGGCTCCATCATCACCCC 863
QY 2006 TGAAGTGGTGTCTCTGCTGAGCTCAATGCTTTTTCAGGATGACAAAATTTCAAGTACTCAGA 2065
Db 864 CGAGTGGATCGTGACAGCGGCCACTGC-----GTGGAATAAATCTTTAACAATCCATG 917
QY 2066 CTACAGATGTGACGCGCTTCTTGGGTCTGCTGGAGCAGAGCAAGCGCAGTGGCTCTGG 2125
Db 918 GCAT-----TGGACGGCAATTTGGGGGATTTTGA-----CAATCTTTCATGTTCTAT 966
QY 2126 GGTGAGGAGTGAAGCTCAAACTATCATCACCCACCTCTCTTCAATGATTTCACTT 2185
Db 967 GGAGCCGATACCAAG--TAGAAAAGTGAATTTCTTCAATCCAAATATGATCTCAAGACAA 1025
QY 2186 CGACTATGACATCGCTTGTGAGAGCTGGAGAGTGGGTGAGTACAGCAGCCTGCTGG 2245
Db 1026 GAACAATGACATGCGCTGATGAAGCTGCAGAGCCTCTGACTTTCAACGACCTAGTGAA 1085
QY 2246 CCCATCTGCTGCTGATGCTACCCATGTTTCCCTGTGGCAAGGCCATCTGGGTAC 2305
Db 1086 ACCAGTGTGTGCTGCCAAACCCAGGCATGATGCTGCAGCCAGACAGCTCTGCTGATTC 1145
QY 2306 AGGCTGGGGGCACACAAAAGAGGGAGGTA-----CCGAGCGCTGATCTCTGAGAGGG 2359
Db 1146 CGGTGGGGGGCCACCGAGGAGAAAGGAGACCTCAGAGTGTCTGACCGTGCAGAGGT 1205
QY 2360 TGAGATCGGTGTATCAACACGAGCACCTGTGAGGACCTCATGCCGAGCAGATCACCCC 2419
Db 1206 GCTTCTCATTTGAGACACAGAGATGCAACAGCAGATATGTCTATGACAACTGATCACACC 1265
QY 2420 ACNAATGATGTGTGGGTTCCTCAGTGGGGGTGTGACTCTCTGCCAGGCTGACTCTGG 2479
Db 1266 AGCATGATCTGTGCCGCTTCTTGCAGGGGAAACGTCGATTTTSCCAGGGTGAAGTGG 1325
QY 2480 TGGCCCTTGTCAAGCGCGAGAAAGATGGCGAATGTTCCAGGCTGTGTGGTGGAGCTG 2539
Db 1326 AGGCTCTGTGTCTACTTTCGAGAAACAATATCTG---GTGGCTGATAGGGGATACAAGCTG 1382
QY 2540 GGGTGAAGGTGGCTCAGAGGAACAAGCCAGGCGGTGTACAAAGGCTCCCTGTAGTTCG 2599
Db 1383 GGGTCTGGCTGTGCCAAAGCTTACAGCAGGAGTGTACGGGAATGTGATGGTATTTCAC 1442
QY 2600 GGAAGTCAAGAGCAGACACTGGGTATA 2629
Db 1443 GGAAGTCAAGAGCAGACACTGGGTATA 2629

RESULT 14
US-09-016-366A-20
; Sequence 20, Application US/09016366A
; Patent No. 5955431
; GENERAL INFORMATION:
; APPLICANT: Stevens, Richard L.
; APPLICANT: Huang, Chifu
; TITLE OF INVENTION: WAST CELL PROTEASE PEPTIDE
; TITLE OF INVENTION: INHIBITORS
; NUMBER OF SEQUENCES: 65
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wolf, Greenfield & Sacks, P.C.
; STREET: 600 Atlantic Avenue
; CITY: Boston
; STATE: MA
; COUNTRY: U.S.A.
; ZIP: 02210-2211
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/016,366A
; FILING DATE: January 30, 1998

CLASSIFICATION: 530
PRIOR APPLICATION DATA: 60/037,090
FILING DATE: 05-FEB-1997
ATTORNEY/AGENT INFORMATION:
NAME: Plumer, Elizabeth R.
REGISTRATION NUMBER: 36,637
REFERENCE/DOCKET NUMBER: B0801/7093
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-720-3500
TELEFAX: 617-720-2441
TELEX:
INFORMATION FOR SEQ ID NO: 20:
SEQUENCE CHARACTERISTICS:
LENGTH: 1128 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-09-016-366A-20

Query Match 3.3%; Score 101.2; DB 2; Length 1128;
Best Local Similarity 53.1%; Pred. No. 3.1e-17;
Matches 394; Conservative 0; Mismatches 288; Indels 60; Gaps 6;

Qy 1903 GCGTGGTGGTGGACGAATCGGACGAGGCGGAGTGGCCCTGGCAGGTGAGCC-----1956
Db 87 GCATCGTTGGGGGTGAGAGGCCCGCCAGGAGCAAGTGGCCCTGGCAGGTGAGCTGAGAG 146

Qy 1957 TCCAGGCCCTGGCCAGGCCACTTGTGTGGGGCTCGCTCATCTCTCTGACTGGCTGG 2016
Db 147 TCCAGGCCCTGATGATGCACTTCTGCGGGGCTCCCTCATCCACCCCGTGGGTGC 206

Qy 2017 TCTCTGAGCTCATTTGCTTTTCAGGATGACAAAATTTCAAGTACTCAGACTACACGATGT 2076
Db 207 TGACCGCAGCGCATCTGGTGGGACCGGACGCTCA-----GG 242

Qy 2077 GGACGGCTTCTGGGTCTGCTGGAACGAGCAAGCGAGTGCCTCTGGGGTGGAGAGC 2136
Db 243 ATCTGGCCGCTCAGGTGCACTGGGGAGAGCAGCTTACTACGAGCAGCTGC 302

Qy 2137 TGAAGCTCAACGATATCATCCACCCCTTCTTCAATGATTTCACTTCACTATGACA 2196
Db 303 TGCCGGTCAGCAGGATCATCTGCAACCCACAGTTTCTACACCGCCAGATCGGAGCGACA 362

Qy 2197 TCGCCTTCTGAGCTGAGAGTGGTGGAGTACAGCAGCGTGTGGCCCGCTCTGCC 2256
Db 363 TCGCCCTCTGAGCTGAGGAGCGGTGAAGGTTCTCCAGCCAGCTCCACACGCTCACCC 422

Qy 2257 TGCCTGATGCTACCCATGTCTTCCCTGTGCAAGGCGCATCTGGGTACAGGCTGGGG-- 2314
Db 423 TGCCCTTCTGCTCAGAGACCTTCCCGCGGGATGCGGTGCTGGGTCACTGGCTGGGGG 482

Qy 2315 ---GCACACAAAGAGGAGTACCGAGCGGTGATCTGCAAGAGGTGAGATCCGCTG 2370
Db 483 ATGTGGCAATGATGAGCGCTCCACCGCCATTTCTCTGAAGCAGGTGAAGTCCCCA 542

Qy 2371 TCATCAACACGACACCTGTCA-----GGACCTCATGCGCCAGCAGATC 2414
Db 543 TAATGGAACCAATTTGTGACGAAATATACACCTTGGCGCTTACCGGGAGACGAGC 602

Qy 2415 ACCCCAGCAATGATGTGT-----GTGGGTTTCTCAGTGGGGGTGTGAGTCTCTGCCAGG 2469
Db 603 TCCGATCTCGGTGACGACATGCTGTGTGCGGGGAACACCCCGAGGAGCTCATGTCCAGG 662

Qy 2470 GTGACTCTGGTGGCCCTTGTCAAGCGCGGAGAAAGATGGCGAATGTTTCAGGCTGGTG 2529
Db 663 CGGACTCCGAGGCGCCCTGTGTGTGCAAGGTGA---ATGGACCTGGGTGCGAGCGGGCG 719

Qy 2530 TGGTGAGCTGGGGTGAAGGCTGCGCTCAGAGCAACAGCGAGGCTGTACACAGGCTCC 2589
Db 720 TGGTCAGCTGGGGGAGGGCTGTGCCAGCCCAACCGGCTTGGCATCTACACCGGTCA 779

Qy 2590 CTGTAGTTGGGACTGGATCAA 2611
Db 780 CCTACTACTTGGACTGGATCCA 801

RESULT 15
US-08-978-404B-15
Sequence 15, Application US/08978404B
Patent No. 5968782
GENERAL INFORMATION:
APPLICANT: Stevens, Richard L.
TITLE OF INVENTION: MAST CELL PROTEASE THAT CLEAVES
NUMBER OF SEQUENCES: 74
CORRESPONDENCE ADDRESS:
ADDRESSEE: Wolf, Greenfield & Sacks, P.C.
STREET: 600 Atlantic Avenue
CITY: Boston
STATE: MA
COUNTRY: U.S.A.
ZIP: 02210-2211
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/978,404B
FILING DATE: 25-NOV-97
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/032,354
FILING DATE: 04-DEC-1996
ATTORNEY/AGENT INFORMATION:
NAME: Plumer, Elizabeth R.
REGISTRATION NUMBER: 36,637
REFERENCE/DOCKET NUMBER: B0801/7090
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-720-3500
TELEFAX: 617-720-2441
TELEX:
INFORMATION FOR SEQ ID NO: 15:
SEQUENCE CHARACTERISTICS:
LENGTH: 1128 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-978-404B-15

Query Match 3.3%; Score 101.2; DB 2; Length 1128;
Best Local Similarity 53.1%; Pred. No. 3.1e-17;
Matches 394; Conservative 0; Mismatches 288; Indels 60; Gaps 6;

Qy 1903 GCGTGGTGGTGGCAGCAATGCGGACGAGGCGGAGTGGCCCTGGCAGGTGAGCC-----1956
Db 87 GCATCGTTGGGGGTGAGAGGCCCGCCAGGAGCAAGTGGCCCTGGCAGGTGAGCTGAGAG 146

Qy 1957 TCCAGGCCCTGGCCAGGCCACTTGTGTGGGGCTCGCTCATCTCTCTGACTGGCTGG 2016
Db 147 TCCAGGCCCTCATGATGCACTTCTGCGGGGCTCCCTCATCCACCCCGTGGGTGC 206

Qy 2017 TCTCTGAGCTCATTTGCTTTTCAGGATGACAAAATTTCAAGTACTCAGACTACACGATGT 2076
Db 207 TGACCGCAGCGCATCTGGTGGGACCGGACGCTCA-----GG 242

Qy 2077 GGACGGCTTCTGGGTCTGCTGGAACGAGCAAGCGAGTGCCTCTGGGGTGGAGAGC 2136
Db 243 ATCTGGCCGCTCAGGTGCACTGGGGAGAGCAGCTTACTACGAGCAGCTGC 302

Qy 2137 TGAAGCTCAACGATATCATCCACCCCTTCTTCAATGATTTCACTTCACTATGACA 2196
Db 303 TGCCGGTCAGCAGGATCATCTGCAACCCACAGTTTCTACACCGCCAGATCGGAGCGACA 362

Qy 2197 TCGCCTTCTGAGCTGAGAGTGGTGGAGTACAGCAGCGTGTGGCCCGCTCTGCC 2256
Db 363 TCGCCCTCTGAGCTGAGGAGCGGTGAAGGTTCTCCAGCCAGCTCCACACGCTCACCC 422

Qy 2257 TGCCTGATGCTACCCATGTCTTCCCTGTGCAAGGCGCATCTGGGTACAGGCTGGGG-- 2314
Db 423 TGCCCTTCTGCTCAGAGACCTTCCCGCGGGATGCGGTGCTGGGTCACTGGCTGGGGG 482

Qy 2315 ---GCACACAAAGAGGAGTACCGAGCGGTGATCTGCAAGAGGTGAGATCCGCTG 2370
Db 483 ATGTGGCAATGATGAGCGCTCCACCGCCATTTCTCTGAAGCAGGTGAAGTCCCCA 542

Qy 2371 TCATCAACACGACACCTGTCA-----GGACCTCATGCGCCAGCAGATC 2414
Db 543 TAATGGAACCAATTTGTGACGAAATATACACCTTGGCGCTTACCGGGAGACGAGC 602

Qy 2415 ACCCCAGCAATGATGTGT-----GTGGGTTTCTCAGTGGGGGTGTGAGTCTCTGCCAGG 2469
Db 603 TCCGATCTCGGTGACGACATGCTGTGTGCGGGGAACACCCCGAGGAGCTCATGTCCAGG 662

Qy 2470 GTGACTCTGGTGGCCCTTGTCAAGCGCGGAGAAAGATGGCGAATGTTTCAGGCTGGTG 2529
Db 663 CGGACTCCGAGGCGCCCTGTGTGTGCAAGGTGA---ATGGACCTGGGTGCGAGCGGGCG 719

Qy 2530 TGGTGAGCTGGGGTGAAGGCTGCGCTCAGAGCAACAGCGAGGCTGTACACAGGCTCC 2589
Db 720 TGGTCAGCTGGGGGAGGGCTGTGCCAGCCCAACCGGCTTGGCATCTACACCGGTCA 779

QY 2197 TCGCCTTGTGAGTGGAGAGTGGTGGAGTACAGCACCGTGTGGGCCCCCATCTGCCC 2256
Db |||||
QY 363 TCGCCTTGTGAGTGGAGAGCGGTGAAGGTCTCCAGCCACGTCCACACGGTCAACC 422
Db |||||
QY 2257 TGCCTGATGCTACCCATGCTTCCCTGTGCAAGGCCCATCTGGGTCAAGGCTGGG-- 2314
Db |||||
QY 423 TGCCTTGCCTCAGAGACCTTCCCTCCGGGATGCCGTGGGTCACTGGCTGGGGCG 482
Db |||||
QY 2315 ----GCACACAAAGAGGAGGTACCGGAGCGCTGATCCTGCAAGGGTGAGATCCCTG 2370
Db |||||
QY 483 ATGTGGACAAATGATGAGCGCCTCCACCGCCATTTCTCTGAAGCAGGTGAAGTCCCCA 542
Db |||||
QY 2371 TCATCAACGAGACCACTGTGA-----GGACCTCATGCCGAGCAGATC 2414
Db |||||
QY 543 TAATGGAACCCACATTTGTGACGCAAAATACCACTTGGCGCCTACACGGGAGACGACG 602
Db |||||
QY 2415 ACCCCACGAATGATGTG-----GTGGGTTTCTCAGTGGGGTGTGGACTCTGCCAGG 2469
Db |||||
QY 603 TCCGCATGTCCTGACGACATGTGTGTGCCGGGAACACCGGAGGACTCATGCCAGG 662
Db |||||
QY 2470 GTGACTGTGGTGGCCCCCTGTCAAGCGGGAGAAAGATGGCGAATTTCCAGGCTGGTG 2529
Db |||||
QY 663 GCGACTCCGGAGGGCCCCCTGGTGTCAAGGTGA---ATGGCACCTGGCTGCAGGGGGCG 719
Db |||||
QY 2530 TGGTGAGCTGGGGTGAAGGCTGCGCTCAGAGGAACAAGCCAGGCGGTGTACACAGGCTCC 2589
Db |||||
QY 720 TGGTCACTGGGGGAGGGCTGTGCCCAAGCCCAACCGGCTGGCATCTACACCCGTTGCA 779
Db |||||
QY 2590 CTGTAGTTCCGGACTGGATCAA 2611
Db |||||
QY 780 CTTACTTGGACTGGATCCA 801
Db |||||

Search completed: February 21, 2004, 07:15:18
Job time : 193 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 21, 2004, 05:28:33 ; Search time 1052 Seconds
(without alignments)
10875.821 Million cell updates/sec

Title: US-09-900-751-1

Perfect score: 3106

Sequence: 1 catggtagacggctgcgcgg.....ttaaaaaaaaaaaaaa 3106

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 2449703 seqs, 1841816367 residues

Total number of hits satisfying chosen parameters: 4899406

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA:*

1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq.*
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq.*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq.*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq.*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq.*
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq.*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq.*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq.*
9: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq.*
10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq.*
11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq.*
12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq.*
13: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq.*
14: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq.*
15: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq.*
16: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq.*
17: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq.*
18: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	3106	100.0	3106	9	US-09-900-751-1	Sequence 1, Appli
2	1889.8	60.8	3149	12	US-10-295-027-968	Sequence 968, Appl
3	1883.2	60.6	3147	11	US-09-776-191-1	Sequence 1, Appli
4	1883.2	60.6	3147	11	US-09-776-191-49	Sequence 49, Appl
5	1883.2	60.6	3147	12	US-10-147-211A-1	Sequence 1, Appli
6	1883.2	60.6	3147	12	US-10-147-211A-3	Sequence 3, Appli
7	1883.2	60.6	3147	12	US-10-156-214A-1	Sequence 1, Appli
8	1883.2	60.6	3147	13	US-10-190-030B-1	Sequence 1, Appli
9	1883.2	60.6	3147	13	US-10-190-030B-3	Sequence 3, Appli
10	1883.2	60.6	3147	13	US-10-267-219-1	Sequence 1, Appli
11	1883.2	60.6	3147	13	US-10-267-219-3	Sequence 3, Appli
12	1883.2	60.6	3147	13	US-10-302-840A-1	Sequence 1, Appli
13	1883.2	60.6	3147	13	US-10-302-840A-3	Sequence 3, Appli
14	1883.2	60.6	3147	13	US-10-112-221A-1	Sequence 1, Appli
15	1883.2	60.6	3147	13	US-10-112-221A-3	Sequence 3, Appli

16	1883.2	60.6	3147	13	US-10-104-271-1	Sequence 1, Appli
17	1883.2	60.6	3147	13	US-10-104-271-3	Sequence 3, Appli
18	1883.2	60.6	3147	15	US-10-099-700A-1	Sequence 1, Appli
19	1883.2	60.6	3147	15	US-10-099-700A-3	Sequence 3, Appli
20	1559.2	50.2	2756	9	US-09-925-301-351	Sequence 351, App
21	1532.4	49.3	3112	15	US-10-097-340-311	Sequence 311, App
22	556.4	17.9	726	15	US-10-092-004A-1	Sequence 1, Appli
23	328	10.6	472	11	US-09-918-995-35415	Sequence 35415, A
24	306.8	9.9	434	10	US-09-736-457-1480	Sequence 1480, Ap
25	306.8	9.9	434	10	US-09-902-941-1480	Sequence 1480, Ap
26	306.8	9.9	434	10	US-09-849-626-1480	Sequence 1480, Ap
27	306.8	9.9	434	13	US-10-113-872-1480	Sequence 1480, Ap
28	306.8	9.9	434	15	US-10-017-754-1480	Sequence 1480, Ap
29	302.6	9.7	429	10	US-09-867-701-5310	Sequence 5310, Ap
30	265	8.5	591	9	US-09-815-343-1503	Sequence 1503, Ap
31	246.8	7.9	393	10	US-09-960-352-5954	Sequence 5954, Ap
32	240.8	7.8	2672	11	US-09-776-191-9	Sequence 9, Appli
33	240.8	7.8	2672	12	US-10-156-214A-9	Sequence 9, Appli
34	240.8	7.8	3104	11	US-09-776-191-7	Sequence 7, Appli
35	240.8	7.8	3104	12	US-10-156-214A-7	Sequence 7, Appli
36	239.2	7.7	1327	10	US-09-978-295A-170	Sequence 170, App
37	239.2	7.7	1327	10	US-09-978-697-170	Sequence 170, App
38	239.2	7.7	1327	10	US-09-978-192A-170	Sequence 170, App
39	239.2	7.7	1327	10	US-09-999-832A-170	Sequence 170, App
40	239.2	7.7	1327	11	US-09-978-189-170	Sequence 170, App
41	239.2	7.7	1327	11	US-09-978-608A-170	Sequence 170, App
42	239.2	7.7	1327	11	US-09-978-585A-170	Sequence 170, App
43	239.2	7.7	1327	11	US-09-978-191A-170	Sequence 170, App
44	239.2	7.7	1327	11	US-09-978-403A-170	Sequence 170, App
45	239.2	7.7	1327	11	US-09-978-564A-170	Sequence 170, App

ALIGNMENTS

RESULT 1
US-09-900-751-1
; Sequence 1, Application US/09900751
; Patent No. US20020026653A1
; GENERAL INFORMATION:
; APPLICANT: Allen, Keith D
; TITLE OF INVENTION: TRANSGENIC MICE CONTAINING SERINE
; TITLE OF INVENTION: PROTEASE GENE DISRUPTIONS
; FILE REFERENCE: R-386
; CURRENT APPLICATION NUMBER: US/09/900,751
; PRIOR FILING DATE: 2001-07-06
; PRIOR APPLICATION NUMBER: US 60/217,449
; PRIOR FILING DATE: 2000-07-10
; PRIOR APPLICATION NUMBER: US 60/223,170
; PRIOR FILING DATE: 2000-08-07
; PRIOR APPLICATION NUMBER: US 60/223,460
; PRIOR FILING DATE: 2000-08-07
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 3106
; TYPE: DNA
; ORGANISM: Mus musculus
US-09-900-751-1

Query Match 100.0%; Score 3106; DB 9; Length 3106;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 3106; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CATGGTAGACGGCTGCCCGGAGGACCAACGCTGTGAGACCGCGATCGGACCGCCAAA 60

Db 1 CATGGTAGACGGCTGCCCGGAGGACCAACGCTGTGAGACCGCGATCGGACCGCCAAA 60

Qy 61 CCATGGTAGCAATCGGGGCGGAGCCGAGGGGCTCTCAGGACTTCGGCGCGGAC 120

Db 61 CCATGGTAGCAATCGGGGCGGAGCCGAGGGGCTCTCAGGACTTCGGCGCGGAC 120

2281 CTGCTGGAAGCCCATCTGGGTACAGGCTGGGGGCACACAAAAGAGGAGGTACCGGAG 2340
2341 CGCTGATCTCGAGAAGGGTCAGATCGTGTATCAACACAGACACACCTGTGAGACCTCA 2400
2341 CGCTGATCTCGAGAAGGGTCAGATCGTGTATCAACACAGACACACCTGTGAGACCTCA 2400
2401 TGCCGACGAGATCACCCACGAATGATGTGTGGGTTCCTCAGTGGGGGTGTGACT 2460
2401 TGCCGACGAGATCACCCACGAATGATGTGTGGGTTCCTCAGTGGGGGTGTGACT 2460
2461 CTTGCCAGGGTACTCTGTGTGGGCTTGTCAAGCGCGGAGAAAGATGGCGAATGTTC 2520
2461 CTTGCCAGGGTACTCTGTGTGGGCTTGTCAAGCGCGGAGAAAGATGGCGAATGTTC 2520
2521 AGGCTGTGTGTGAGTGGGTGAGGCTGAGGCTGCGCTCAGAGGAAACAGCCAGGCTGTACA 2580
2521 AGGCTGTGTGTGAGTGGGTGAGGCTGCGCTCAGAGGAAACAGCCAGGCTGTACA 2580
2581 CAAGGCTCCCTGTAGTTCGGGACTGGATCAAGAGACACTGGGGTATAGCAGCATGGAC 2640
2581 CAAGGCTCCCTGTAGTTCGGGACTGGATCAAGAGACACTGGGGTATAGCAGCATGGAC 2640
2641 AGACAGCGGACCAACACACAGGATGCCCGACATGCACACCTGTGATACAGGAGAG 2700
2641 AGACAGCGGACCAACACACAGGATGCCCGACATGCACACCTGTGATACAGGAGAG 2700
2701 GAACACTGACGACATTTATGCTGTGGGCTCCCGCCCAACACACACAGCAGTGTGAAC 2760
2701 GAACACTGACGACATTTATGCTGTGGGCTCCCGCCCAACACACACAGCAGTGTGAAC 2760
2761 GCATCTTAGGACTCAGAGTTCCTCAAGTGGGACCCCTCAAGAGTTCGAGAGAGAAC 2820
2761 GCATCTTAGGACTCAGAGTTCCTCAAGTGGGACCCCTCAAGAGTTCGAGAGAGAAC 2820
2821 TTGCGTCTAGCGGCCAGCTGGGGCAGAGGTTTGAATGGCAGCCTTCCCTCTAGCC 2880
2821 TTGCGTCTAGCGGCCAGCTGGGGCAGAGGTTTGAATGGCAGCCTTCCCTCTAGCC 2880
2881 CTGAGCTGGGTGAAGATGATGCTGTCCCGGAGAGTCTCTTCCAACTGTCAATGAGCTCCC 2940
2881 CTGAGCTGGGTGAAGATGATGCTGTCCCGGAGAGTCTCTTCCAACTGTCAATGAGCTCCC 2940
2941 GGGAGCCTATGGGAGAGGGGTCAAGGTCATCTTTTTCAGGAAGCGCCAGCCCTAGGA 3000
2941 GGGAGCCTATGGGAGAGGGGTCAAGGTCATCTTTTTCAGGAAGCGCCAGCCCTAGGA 3000
3001 ACCCGAGAAAGAGTGTACTAAGCTGAATGTTTTCGCTGTGCGAGGGTGGGTAT 3060
3001 ACCCGAGAAAGAGTGTACTAAGCTGAATGTTTTCGCTGTGCGAGGGTGGGTAT 3060
3061 TTGAGAGTAAACATTTTATTTCTTTTAAAAAATAAAAAAAAAAAAA 3106
3061 TTGAGAGTAAACATTTTATTTCTTTTAAAAAATAAAAAAAAAAAAA 3106

RESULT 2

US-10-295-027-968
; Sequence 968, Application US/10295027
; Publication No. US2003023250A1
; GENERAL INFORMATION:
; APPLICANT: Afar, Daniel
; APPLICANT: Aziz, Natasha
; APPLICANT: Ginsberg, Wendy M.
; APPLICANT: Gish, Kurt C.
; APPLICANT: Glynn, Richard
; APPLICANT: Hevezi, Peter A.
; APPLICANT: Mack, David H.
; APPLICANT: Murray, Richard
; APPLICANT: Watson, Susan R.
; APPLICANT: Eos Biotechnology, Inc.
; TITLE OF INVENTION: Methods of Diagnosis of Cancer, Compositions and
; TITLE OF INVENTION: Methods of Screening for Modulators of Cancer
; FILE REFERENCE: 018501-012500US

; CURRENT APPLICATION NUMBER: US/10/295,027
; CURRENT FILING DATE: 2002-11-13
; PRIOR APPLICATION NUMBER: US 09/663,733
; PRIOR FILING DATE: 2000-09-15
; PRIOR APPLICATION NUMBER: US 60/350,666
; PRIOR FILING DATE: 2001-11-13
; PRIOR APPLICATION NUMBER: US 60/335,394
; PRIOR FILING DATE: 2001-11-15
; PRIOR APPLICATION NUMBER: US 60/332,464
; PRIOR FILING DATE: 2001-11-21
; PRIOR APPLICATION NUMBER: US 60/334,393
; PRIOR FILING DATE: 2001-11-29
; PRIOR APPLICATION NUMBER: US 60/340,376
; PRIOR FILING DATE: 2001-12-14
; PRIOR APPLICATION NUMBER: US 60/347,211
; PRIOR FILING DATE: 2002-01-08
; PRIOR APPLICATION NUMBER: US 60/347,349
; PRIOR FILING DATE: 2002-01-10
; PRIOR APPLICATION NUMBER: US 60/355,250
; PRIOR FILING DATE: 2002-02-08
; PRIOR APPLICATION NUMBER: US 60/356,714
; PRIOR FILING DATE: 2002-02-13
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 1386
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 968
; LENGTH: 3149
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-10-295-027-968

Query Match 60.8%; Score 1889,8; DB 12; Length 3149;
Best Local Similarity 81.4%; Pred.No. 0;
Matches 2222; Conservative 0; Mismatches 504; Indels 5; Gaps 3;
QY 50 GACCGCCAAACCATGGGTAGCAATCGGGGCCCAAGGCCGAGGGGGCTCTCAGGACTT 109
DB 23 GGCCTCGGGACCATGGGAGCGATCGGCCCGCAAGGGCGGAGGGGCCCGAAGGACTT 82
QY 110 CGCGCGGGGACTCAAGTACAACTCCCGGGCTAGAGAACATGAATGGCTTTGAGAGGGTGT 169
DB 83 CGCGCGGGGACTCAAGTACAACTCCCGGCCACGAGAAAGTGAATGGCTTTGAGAGGCGGT 142
QY 170 GGAATTCCTCGCTCGCAACAAATGCCAAGAAAGTGGAGAGCGAGGCCGCCAGGCGCTGGGT 229
DB 143 GGAATTCCTCGCTCGCAACAAATGCCAAGAAAGTGGAGAGCGAGGCCGCCAGGCGCTGGGT 202
QY 230 GGTGCTGGTGGCAGTGTCTGTTCAGCTTCCTCTTGTCTCTCCCTCATGGCTGGCTTGTCTGT 289
DB 203 GGTGCTGGCAGCGCTGTCTGATCGGCTCTCTTGTCTTGTCTGGGATGGCTTCTCTGT 262
QY 290 GTGGCACTTCCATTATCGGAATGTGGGGTTCAGAAAGTCTTCAATGGCCATCTGAGGAT 349
DB 263 GTGGCACTTTCAGATACCGGGAGCGTGGTGTCCAGAAAGGCTTCAATGGCTTACATGAGGAT 322
QY 350 CACAAATGAGATCTTTCTGATGCGGTATGAGAACTCCACTCCACAGAGTTTATCAGCCT 409
DB 323 CACAAATGAGATTTTGTGATGCGCTTACGAGAACTCCAACTCACTCAGTTTGTAAAGCCT 382
QY 410 GGCACGCCAGGTGAAGGAGGCGCTGAAGCTGTGTACAATGAAGTCCCTGTCTCGGGTCC 469
DB 383 GGCACGAAAGGTGAAGGAGCGCTGAAGCTGTGTACAGCGGAGTCCCATTCCTGGGCC 442
QY 470 CTACACAGAGAGTGGCTGTAACTGCCCTTCAGTGGGCGAGTGTATCGCTTACTCTG 529
DB 443 CTACACAGAGGAGTGGCTGTGAAGCGCTTCAGCGGCTTCAGCGAGGAGCGCTCATCGCTTACTCTG 502
QY 530 GTCAGAGTTTCAAGATCCCGCCCAACCTGCGCAGAGAGGTTGATCGCGCATGCTGTGGA 589
DB 503 GTCTGAGTTTCAAGATCCCGCCCAACCTGCGCAGAGGAGGCGCGCTCATGCGCCAGGA 562
QY 590 GCGAGTTGTAAATTCGCCACCCCGAGCAGCGGCACTGAAATCTCTCGTGTAACTCTGT 649

RESULT 3
US-09-776-191-1
; Sequence 1, Application US/09776191
; Publication No. US20030119168A1
; GENERAL INFORMATION:
; APPLICANT: Edwin L. Madison
; APPLICANT: Edgar O. Ong
; APPLICANT: Jiumn-Chern Yeh
; APPLICANT: Corvas International, Inc.
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING
; TITLE OF INVENTION: TRANSMEMBRANE SERINE PROTEASES, THE ENCODED PROTEINS AND
; TITLE OF INVENTION: METHODS BASED THEREON
; FILE REFERENCE: 24745-1607
; CURRENT APPLICATION NUMBER: US/09/776,191
; CURRENT FILING DATE: 2001-02-02
; PRIOR APPLICATION NUMBER: 60/213,124
; PRIOR FILING DATE: 2000-06-22
; PRIOR APPLICATION NUMBER: 60/234,840
; PRIOR FILING DATE: 2000-06-22
; PRIOR APPLICATION NUMBER: 60/179,982
; PRIOR FILING DATE: 2000-02-03
; PRIOR APPLICATION NUMBER: 60/183,542
; PRIOR FILING DATE: 2000-02-18
; PRIOR APPLICATION NUMBER: 09/657,968
; PRIOR FILING DATE: 2000-02-08
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 3147
; TYPE: DNA
; ORGANISM: Homo Sapien
; FEATURE:
; OTHER INFORMATION: Nucleotide encoding MTSP1
; NAME/KEY: CDS
; LOCATION: (23)...(2589)
; PUBLICATION INFORMATION:
; AUTHORS: O'Brien, T.J. and Tanimoto, H.
; DATABASE ACCESSION NUMBER: GenBank AR081724
; PATENT DOCUMENT NUMBER: US Pat 5972616
; PATENT FILING DATE: 1998-02-20
; PUBLICATION DATE: 1999-10-26
US-09-776-191-1
Query Match 60.6%; Score 1883.2; DB 11; Length 3147;
Best Local Similarity 81.2%; Pred. No. 0;
Matches 2223; Conservative 0; Mismatches 508; Indels 5; Gaps 3;
QY 45 GATCGACCGCCAAACCAATGGGTAGGATCGGGCCGCAAGCCGCGAGGGGGCTCTCAG 104
Db 5 GAGCGGCTCGGGGTACCAATGGGAGCGATCGGGCCGCAAGGGCGGAGGGGGCCGGAAG 64
QY 105 GACTTCGGCGCGGACTCAAGTACAACTCCCGCTAGAGAACATGAATGGCTTTGAGGAG 164
Db 65 GACTTCGGCGCGGACTCAAGTACAACTCCCGCTAGAGAACATGAATGGCTTTGAGGAG 124
QY 165 GGTGTGAGTTCCTGCTCGAACAAATCCAGAAAGTGGAGAGCGAGGCCCCAGGGCGC 224
Db 125 GCGGTGGAGTTCCTGCGAGTCAACACGCTCAAGAGGTGGAAGCATGGCCCGCGGGCGC 184
QY 225 TGGGTGCTGCTGGTGGCAGTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 284
Db 185 TGGGTGCTGCTGGCAGCGGTGTGATCGGCTCCTCTGCTGCTGCTGCTGCTGCTGCTGCT 244
QY 285 CTGGTGTGCACTTCCATATCGGAATGCGGGTTCAAAAAGTCTTCAATGGCCATCTG 344
Db 245 CTGGTGTGCACTTCCAGTACCGGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 304
QY 345 AGATACAAAATGAGATCTTTCTGGATCGGTATGAGAACTCCACCTCCACAGAGTTTATC 404
Db 305 AGATACAAAATGAGAAATTTTGTGATGCTTACGAGAACTCCAACTCCCACTGAGTTTGA 364
QY 405 AGCCTGGCCAGCCAGGTGAGGAGCGGCTGAAGTCTGTACAAAGTCCCTGTCTCTG 464

Db 365 AGCCTGGCCAGCAAGGTGAGGAGCGGCTGAAGTCTGTATCAGCGGATCCCATTCCTG 424
QY 465 GGTTCCTTACCAACAAGAGTCGGCTGTAACTGCTTCAAGTGAAGGCGAGTGTATCGCTAC 524
Db 425 GGGCCCTTACCAACAAGAGTCGGCTGTGAAGGCTTTCAAGGAGGCGAGCTCATCGCTAC 484
QY 525 TACTGTGTCAGAGTTTCAAGATCCCGCCACACCTGGGAGAGAGGTTGATCGCGCATGGCT 584
Db 485 TACTGTGTCAGATTTAGCATCCCGACACCTGGTGGAGAGGCGGCTCATGGCC 544
QY 585 GTGAGCGAGTTTGAACATTTGCCACCCGAGCAGCGGCACTGAAATCTCTTCTGCTTAA 644
Db 545 GAGGAGCGGTAGTATGCTGCTGCGCGCGGCGCTCTCTTCTGAGTCTCTTGTGGTCA 604
QY 645 TCTGTGTGGCTTCCCGCATTTGACCCAGATGCTGCGAGGACTCAGAGCAACAGCTGC 704
Db 605 TCAGTGTGGCTTCCCGCATTTCCCGCATTTCCAAACAGTACAGAGGAGCCAGCAACAG 664
QY 705 AGTTTTGGCCCTCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 764
Db 665 AGCTTTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 724
QY 765 AACAGTCCCTTACCCGCGCATGCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 824
Db 725 GACAGCCCTTACCCGCTCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 784
QY 825 GTGCTGAGCTTACCTTCCGAAGCTTTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTG 884
Db 785 GTGCTGAGCTTACCTTCCGCGATTTGACCTTGGCTTCTGCGAGCGGCGGCGAGC 844
QY 885 CTGGTCACTGCTGATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 944
Db 845 CTGGTGAAGTGTACAACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 904
QY 945 GGCACTTCTTACCTTCTTACAACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1004
Db 905 GGCACCTTACCTTCTTACAACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 964
QY 1005 AGCTGATACCAATCTGACCGGCGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1064
Db 965 ACAGTGAACCACTGAGCGGCGCATCCGCTTGTGAGGCGCACCTTCTTCCAGCTG 1024
QY 1065 CCCAAGATGAGCAGCTGTGGCGGCTTTTGTAGTGACACCCAGGAGCATTTAGCAGCCCC 1124
Db 1025 CTTAGATGAGCAGCTGTGGAGCGCTTACGTAAGCCCGAGGAGCATTTCAAGCCCC 1084
QY 1125 TACTATCCAGGCGCATCCCGCCCAACATCACTGCAATGGAATATCAAGTGTGCCAAC 1184
Db 1085 TACTATCCAGGCGCATCCCGCCCAACATCACTGCAATGGAATATCAAGTGTGCCAAC 1144
QY 1185 AACCGGAACGTGAAGTGTGCTTCAAACTCTTCTATGCTGCTGCTGCTGCTGCTGCTG 1244
Db 1145 AACCGAGATGTGAAGTGTGCTTCAAACTCTTCTATGCTGCTGCTGCTGCTGCTGCT 1204
QY 1245 GGTCTCTGACCAAGGACTATGTGAGATCAACGGGAGAGTACTGCGGTGAGAGTCC 1304
Db 1205 GGCACCTGCCCCAAGGACTATGTGAGATCAATGGGAGAAATCTGCGGAGAGAGTCC 1264
QY 1305 CAGTTTGTGTGAGCAGCAACAGCAGCAAGATTAAGTCCACTTCCATTTGATCACTCG 1364
Db 1265 CAGTTTGTGTGAGCAGCAACAGCAGCAAGATTAAGTCCACTTCCATTTGATCACTCG 1324
QY 1365 TACAGGACACCGGTTCTTCTAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1424
Db 1325 TACAGGACACCGGTTCTTCTAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1384
QY 1425 GGGATGTTTCTGTCAGAGCTGGACGGTGCATCCGAAAGAACTGCGCTGCGAGCGGTGG 1484
Db 1385 GGGCAGTTTCACTGCGCGCAGCGGGCGGTGATCCGGAAGGAGCTGCGCTGTGATGGTGG 1444
QY 1485 GCAGCTGCCCGGATTTAGTGTGAGGCTTACTGCGCGATGCAATGCCACCCAGCTTC 1544

Db 1445 GCCAGCTGCACCGACACAGCAGTACAGTCACTCACTGAGTTCGACGCCGCCACCAAGTTC 1504
Qy 1545 ACCTGCAAAAACCAAGTTCTGCAAGCCCTCTCTTGGGTCTGTGACAGTGTCAACAGACTGT 1604
Db 1505 ACCTGCAAGAACAAAGTTCTGCAAGCCCTCTCTTGGGTCTGTGACAGTGTGAAAGACTGC 1564
Qy 1605 GGGACGGAAGTGTGACGAGGAGGCTGACGTGTCTGTCTGGAGTTCCTCAAGTGTCCAAAT 1664
Db 1565 GGAGACAAACAGCAGCAGCAGGAGTGTCTGCGGCCAGACTTCAGGTGTTCCTCAAT 1624
Qy 1665 GGGAAAGTGTCTCTCCCTCAGAGCCAGAGTGTAAATGGGAAGGACCACTGTGGAGATGGGTCT 1724
Db 1625 GGGAAAGTGTCTCTCGAAAGCCAGCAGTGTCAATGGGAAGGACCACTGTGGGAAGCGGTCTC 1684
Qy 1725 GACGAGGCTTCATGTGACAGCGTGAATGTCTCTTGTCCACCAATATATACCTACCGCTGC 1784
Db 1685 GACGAGGCTCTCTGCCCCAAGGTGAACCTGTCTCTTGTACCAACACACCTACCGCTGC 1744
Qy 1785 CAATATGCTCTCTGTGACGAAGGCAACCTCAGTGTGTATGGGAAGACGGACTGTAGC 1844
Db 1745 CTCATATGGCTCTCTGTGACGAAGGCAACCTCAGTGTGTACGGGAAGGAGACTGTAGC 1804
Qy 1845 GATCGCTCCGATGAGAAAACTGTGACTGTGGGTGGATCTCTTTACCAACACAGGCTCGC 1904
Db 1805 GACGCTCAGATGAGAGGACTGTGACTGTGGGTGGGTCACTTTCAGACAGACGCTCGT 1864
Qy 1905 GTGTTGGTGTGACGAAGTGTGACGAGGCGAGTGTGGCTGTGGAGGTGAGCTCCAGCC 1964
Db 1865 GTTGTGGGGCACCGATGCGGATGAGGGCAGTGGCCCTGGCAGTAAAGCTGCATGCT 1924
Qy 1965 CTGGGCCAGGCGCACTGTGTGGGGCTCGCTCATCTCTCTGACTGCTGGTCTGTGCA 2024
Db 1925 CTGGGCCAGGCGCACTGTGGGGTCTTCCCTCATCTCTCCAACTGGCTGTCTGCTC 1984
Qy 2025 GCTCAATGCTTTCAGGATGACAAAAATTTCAAGTACTCAGACTACAGATGTGACCGGCC 2084
Db 1985 GCACACTGCTACATCATGATGACAGAGGATTCAGGTACTCAGACCCACGCGTGAACGCC 2044
Qy 2085 TTCTTGGGTCTGTGACGAGCAGGAGTGTCTGTGGGTGTGAGGAGTGAAGCTC 2144
Db 2045 TTCTTGGGTCTGTGACGAGCAGGAGCAGGAGCAGGCGCTGTGGGTGTGAGGAGCGCAGCTC 2104
Qy 2145 AAAGTATATCAACCCACCTCTCTTCAATGATTTCACTTCCACTGACTGATCATCGCTTG 2204
Db 2105 AAGCGCATATCTCCACCCCTCTTCAATGATTTCACTTCCACTGACTGATCATCGCTTG 2164
Qy 2205 CTGAGCTGGAGAAAGTGTGGTGTGACAGCAGCAGTGTGGCCCATCTGCTGCTGCTGAT 2264
Db 2165 CTGAGCTGGAGAAACCGCAGAGTACAGCTCCATGTGTGGCCCATCTGCTGCTGCTGAC 2224
Qy 2265 GCTACCATGTTCTTCCCTGTGGCAAGCCATCTGGGTGTGAGGCTGGGGGACACAAAA 2324
Db 2225 GCTTCCCATGTTCTTCCCTGTGGCAAGCCATCTGGGTGTGAGGCTGGGGGACACACAG 2284
Qy 2325 GAGGAGGTACCGGAGCGCTGATCCTGTGAGAGGCTGAGATCCGTGTATCAACACAGACC 2384
Db 2285 TATGAGGCACTGTGGGTGTGATCTGTGAAAGGCTGAGATCCCGTGTATCAACACAGACC 2344
Qy 2385 ACCTGTGAGGACCTCATGCGCAGCAGATCACCCACGAATGATGTGTGGGTTCCTC 2444
Db 2345 ACCTGTGAGAACCTCTCTGCGCAGCAGATCACCGCGCATGATGTGCGTGGGTTCCTC 2404
Qy 2445 AGTGGGGTGTGAGTCTCTGCGAGGCTGACTCTGTGTGGCCCTTGTCAAGCGCGGAGAAA 2504
Db 2405 AGCGCGCGGTGAGTCTCTGCGAGGCTGATTTCCGGGGGACCCCTGTCCAGCGTGGAGCG 2464
Qy 2505 GATGGCGGAATGTTCCAGGCTGTGTGTGAGCTGGGTGTGAGGCTGGCTGTGAGGAAAC 2564
Db 2465 GATGGCGGAATGTTCCAGGCGGTGTGTGTGAGCTGGGTGTGAGGCTGGCTGTGAGGAAAC 2524
Qy 2565 AAGCAGCGGTGTACAAAGGCTCCCTGTGTGTGTGAGTGTGAGGCTGGATCAAGAGCACTGGG 2624
Db 2525 AAGCAGCGGTGTACAAAGGCTCCCTGTGTGTGTGAGTGTGAGGCTGGATCAAGAGCACTGGG 2584

Qy 2625 GTATAGCAGCATGACAGACAGCGACCAAAACACCCACAGGATGCCGACATGCACA 2684
Db 2585 GTATAGGGGCGGG---GCCACCAATGTGTACACTGCGGGCCACCATCGTCCACC 2641
Qy 2685 CTTGATACAGGAGAGAAACATGACGACATTTATGTGTGGCTTCCCCCCCCCAACACA 2744
Db 2642 CCAAGTGTGCAGC-CCTGAGGCTGGAGACTGGACCGCTGACTGCACACAGGCCCC-CCAGA 2699
Qy 2745 ACCCAGACTGTGAACCTGCATCTTTAGGACTCAGAGT 2780
Db 2700 ACATACACTGTGAACCTCAATCTCCAGGGCTCCAAAT 2735

RESULT 4

US-09-776-191-49
; Sequence 49, Application US/09776191
; Publication No. US20030119169A1
; GENERAL INFORMATION:
; APPLICANT: Edwin L. Madison
; APPLICANT: Edgar O. Ong
; APPLICANT: Jiunn-Chern Yeh
; APPLICANT: Corvas International, Inc.
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING
; TITLE OF INVENTION: TRANSMEMBRANE SERINE PROTEASES, THE ENCODED PROTEINS AND
; FILE OF INVENTION: METHODS BASED THEREON
; FILE REFERENCE: 24745-1607
; CURRENT APPLICATION NUMBER: US/09/776,191
; PRIOR FILING DATE: 2001-02-02
; PRIOR APPLICATION NUMBER: 60/213,124
; PRIOR FILING DATE: 2000-06-22
; PRIOR APPLICATION NUMBER: 60/234,840
; PRIOR FILING DATE: 2000-06-22
; PRIOR APPLICATION NUMBER: 60/179,982
; PRIOR FILING DATE: 2000-02-03
; PRIOR APPLICATION NUMBER: 60/193,542
; PRIOR FILING DATE: 2000-02-18
; PRIOR APPLICATION NUMBER: 09/657,968
; PRIOR FILING DATE: 2000-02-08
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 49
; LENGTH: 3147
; TYPE: DNA
; ORGANISM: Homo Sapien
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1865)...(2590)
; OTHER INFORMATION: Nucleic acid sequence of protease domain of MTPS1
US-09-776-191-49

Query Match 60.6%; Score 1883.2; DB 11; Length 3147;
Best Local Similarity 81.2%; Pred. No. 0;
Matches 2223; Conservative 0; Mismatches 508; Indels 5; Gaps 3;
Qy 45 GATCGGACCCGCAAAACCATGGGTAGCATCGGGGCGCAAGCCGAGGGGCTCTCAG 104
Db 5 GAGCGGCTCTGGGGTACCATGGGGAGCGATCGGGCCCGCAAGGGCGAGGGGCGCCGAAG 64
Qy 105 GACTTCGGCGCGGAGCTCAAGTACAACTCCCGGCTAGAGAACATGAATGGCTTTGAGGAG 164
Db 65 GACTTCGGCGCGGAGCTCAAGTACAACTCCCGGCTAGAGAACATGAATGGCTTTGAGGAA 124
Qy 165 GGTGTGGAGTCTCTGCTCGCAACAATGCGAAGAGTGGAGAGGAGGAGGAGGAGGAGGAG 224
Db 125 GCGTGGAGTCTCTGCGGAGTCAACAGTCAAGAGTGGAGAGGAGGAGGAGGAGGAGGAG 184
Qy 225 TGGGTGTGTGTGTGGAGT 284
Db 185 TGGGT 244
Qy 285 CTGCTGTGGCACTTCCATATATCGAATGTGCGGTTCAAAAAGCTTCTCATGTGCACTCTG 344

Db 245 CTGGTGTGGCAATTTGGAGTACCGGGAGCGTGTGTGTCAGAGAGGTCTTCAATGGCTACATG 304
QY 345 AGGATCACAAATGAGATCTTTCTGGATGCGTATGAGAACTCCACCTCCACAGATTTATC 404
Db 305 AGGATCACAAATGAGAAATTTTGTGATGCTTACGAGAACTCCAACTCCACTGAGTTTGA 364
QY 405 AGCTTGGCCAGGAGTGAAGAGGCGCTGAGAGCTGTCTGTATAAATGAAGTCTCCTGTCTG 464
Db 365 AGCTTGGCCAGGAGTGAAGAGGCGCTGAGAGCTGTCTGTATAAAGGAGTCCCAATTCCTG 424
QY 465 GGTCTCCTACCAAGAAAGTGTGCTGTAATGCTGCTTCAAGTGAAGGAGTGTATCCGCTAC 524
Db 425 GCGCCCTACCAAGAGATGCTGCTGTGACGGCTTTCAGGAGGAGCGTGTATCGCTAC 484
QY 525 TACTGTTCAGATTCAGATCCCGCCACACCTGGCAGAGAGGTTGATCGCGCCATGGCT 584
Db 485 TACTGTCTGAGTTCAGATCCCGCAGCAGCTGTGTGAGAGGAGCGCGCTCATGGCC 544
QY 585 GTGGAGCGAGTTGTAAATTTGCAATTTGCAATTTGCAATTTGCAATTTGCAATTTGCAAT 644
Db 545 GAGGAGCGGTAGTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 604
QY 645 TCTGTGTGGCTTCCCATTCACCCAGAAATGCTGAGAGGAGTCAAGAGAGTGTGATCGCGCCATGGCT 704
Db 605 TCAGTGTGGCTTTCCCGACGAGCTTCCAAACAGTACAGAGAGAGCGAGAGAGAGTGTG 664
QY 705 AGTTTGGCCCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 764
Db 665 AGCTTGGCTGAGAGCGGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGT 724
QY 765 AACAGTCCCTACCGCGCATGCGCTGCGAGTGGTCTGCGGGGAGCGCCAGTCT 824
Db 725 GACAGCCCTACCGCGCATGCGCTGCGAGTGGTCTGCGGGGAGCGCCAGTCT 784
QY 825 GTGCTGAGCTCACCTTCGAAAGCTTTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 884
Db 785 GTGCTGAGCTCACCTTCGAGAGCTTTGACTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 844
QY 895 CTGGTACCGTGTATGATGCTGAGCGCCATGGAACCCAGCTGTGTGCTGCTGCTGCTGCTGCT 944
Db 845 CTGGTACCGTGTATGATGCTGAGCGCCATGGAACCCAGCTGTGTGCTGCTGCTGCTGCTGCT 904
QY 945 GGCACCTTCTACCTCTCTACACCTGATCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1004
Db 905 GGCACCTTCT 964
QY 1005 AGCTGATAACCAATCTGACCGCGCATCTCTGCTTGTGAGGCACTTTCTTCTCAGCTG 1064
Db 965 ACCTGATAACCAACTGAGCGCGCATCTCTGCTTGTGAGGCACTTTCTTCTCAGCTG 1024
QY 1065 CCCAAGATGAGAGCTGTGGCGGCTTTTGTGATGACACCCAGGAGACATTTAGCAGCCCC 1124
Db 1025 CTTAGATGAGAGCTGTGGAGCGCTTACGTAAGGCCAGGAGACATTTCAACAGCCCC 1084
QY 1125 TACTATCCAGGCACTTACCGGCCAATCTCACTGCAATGATGATGATGATGATGATGATGATGATGAT 1184
Db 1085 TACTATCCAGGCACTTACCGGCCAATCTCACTGCAATGATGATGATGATGATGATGATGATGATGAT 1144
QY 1185 AACCGGAACGTGAAGTGTGGCTTTCAAACTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1244
Db 1145 AACCGGAACGTGAAGTGTGGCTTTCAAACTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1204
QY 1245 GCTCTCTGACCAAGGATGATGAGATGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1304
Db 1205 GGCACCTGCGGAGAGCTGAGTGTGAGATGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1264
QY 1305 CAGTTTGTGGTGAAG 1364
Db 1265 CAGTTTGTGGTGAAG 1324
QY 1365 TACAGGAG 1424
Db 1325 TACAGGAG 1384

QY 1425 GGGATGTTTCATGTGCAAGACTGGACGGTGCATCCGAAAGGAAGTGGCTGGCGACGCGCTGG 1484
Db 1385 GGGCAATTTCACTGCGCGACGGGGCGGTGTATCCGGAGAGAGTGGCTGGCTGTGATGGCTGG 1444
QY 1485 GCAGACTCCCGGATTAATGATGATGAGCGTTTACTCCCGATGCAATGCCACCCAGTTTC 1544
Db 1445 GCGCACTCACCGACACAGCGATGAGCTCACTGAGTTGGAGCGCGCGGCCACCGATTC 1504
QY 1545 AGTGCACAAACCAAGTCTGCAAGCGCTCTCTCTGGGTCTGTGACAGTGTCAAGCTGT 1604
Db 1505 AGTGCACAAACCAAGTCTGCAAGCGCTCTCTCTGGGTCTGTGACAGTGTCAAGCTGT 1564
QY 1605 GGGAGCGAAGTGAAGAGGAGGCTGCAAGTGTCTGCTGGGAGTTTCAAGTGTTCGAAT 1664
Db 1565 GGGAGCAACAGGAGGAGGAGGCTGCAAGTGTCTGCGGCCAGACCTTCAGTGTTCGAAT 1624
QY 1665 GGGAGTGTCTCTCTCAGAGCGAGAGTGTATGAGGAAGGACAACTGTGGAGATGGGTCT 1724
Db 1625 GGGAGTGTCTCTCTGAAAGGAGAGTGTCAATGGAAGGAGAGTGTGGGAGCGGTCC 1684
QY 1725 GACGAGGCTTCATGTGACAGGCTGAATGCTCTCTTGCACCAATATACCTACCGCTGC 1784
Db 1685 GACGAGGCTTCATGTGACAGGCTGAATGCTCTCTTGCACCAATATACCTACCGCTGC 1744
QY 1785 CAAATGGCTCTCTGTGAGCAAGGCGCAACCTGAGTGTGATGGGAAGACGAGCTGTAGC 1844
Db 1745 CTAATGGCTCTCTGTGAGCAAGGCGCAACCTGAGTGTGACGGAGAGGAGCTGTAGC 1804
QY 1845 GATGGCTCGATGAGAAACCTGTGCTGTGGCTGCGATCTTTTACCAACAGGCTGCG 1904
Db 1805 GACGCTCAGATGAGAGGACTGTGCGACTGTGGGTGCGGTCAATTCACGAGACAGGCTCGT 1864
QY 1905 GTGGTGTGTGCGACGAATGCGGAGCGGAGTGGCCCTGCGAGGTGAGCTCCACGCG 1964
Db 1865 GTGGTGTGTGCGACGAATGCGGAGCGGAGTGGCCCTGCGAGGTGAGCTCCACGCG 1924
QY 1965 CTGGGCGAGGCGCACTTGTGTGGGCGCTCGCTCATCTCTCTGACTGGCTGGTCTCTGCA 2024
Db 1925 CTGGGCGAGGCGCACTTGTGTGGGCGCTCGCTCATCTCTCTGACTGGCTGGTCTCTGCG 1984
QY 2025 GCTCATGTCTTTCAGATGACAAATTTTCAAGTACTCAGACTACAGATGTGACGCGCC 2084
Db 1985 GCACATGCTACATCGATGACAGAGGATTCAGGTACTCAGACCCCGCAGCTGTGAGCGCC 2044
QY 2085 TTTCTGGTGTGTGAGCAGCAGCAAGCGAGTGGCTGTGGGTGCGAGGAGTGAAGCTC 2144
Db 2045 TTTCTGGTGTGTGAGCAGCAGCAGCAGCGCGCGCCCTGGGGTGCAGGAGCGGAGCTC 2104
QY 2145 AAACGTATCATCACCCACCTTCTTCAATGATTTTCACTTTCGACTATGACATCGCCTTG 2204
Db 2105 AAGCGCATCATCTCCACCCCTTCTTCAATGATTTTCACTTTCGACTATGACATCGCCTTG 2164
QY 2205 CTGGAGCTGGAGAGTGTGGTGAAGTACAGACCTGTGCGCGCCCTGCTGCTGCTGCTGCTGCTGAT 2264
Db 2165 CTGGAGCTGGAGAAACCGGAGAGTACAGCTTCAATGTTGGGCGCCCTGCTGCTGCTGCTGCTG 2224
QY 2265 GCTACCATGCTTCTCTCTGCGCAAGCGCACTCTGGGTGACAGCTGGGGGACACACAAA 2324
Db 2225 GCTTCCCATGCTTCTCTCTGCGGCAAGCGCATCTGGGTGACGCGGTGGGGACACACCGAG 2284
QY 2325 GAGGAGGTACCGAGCGCTGATCTCTGCAAGGAGTGAATTCGATTCATCAACAGAGCTC 2384
Db 2285 TATGGAGCACTGGCGCGCTGATCTTCAAAAGGAGTGAATTCGATTCATCAACAGAGCTC 2344
QY 2385 ACCTGTGAGGAGCTCATCGCGCAGCAGATCAACCCGAGATGATGTGTGGGTTCCTC 2444
Db 2345 ACCTGTGAGGAGCTCATCGCGCAGCAGATCAACCCGAGATGATGTGTGGGTTCCTC 2404
QY 2445 AGTGGGCTGTGAGTCTCTGCGAGGCTGAGTCTGCTGGCGCCCTTGTCAAGCGCGGAGAAA 2504
Db 2405 AGCGGCGGCTGTGAGTCTCTGCGAGGCTGATTCGCGGGGACACCCCTGTTCAGGCTGGAGCG 2464

QY 2505 GATGGGGAATGTTCCAGGCTGGTGTGGTGGGTTGAAGCTGGCTCGCTCAGAGGAAC 2564
Db 2465 GATGGGCGATCTTCCAGGCGGTGTGGTGGGTTGAAGCTGGCTCGCTCAGAGGAAC 2524
QY 2565 AAGCCAGCGGTGACACAAGGCTCCCTGTAGTTGGGACTGGATCAAGAGCACTGGG 2624
Db 2525 AAGCCAGCGGTGACACAAGGCTCCCTGTAGTTGGGACTGGATCAAGAGCACTGGG 2584
QY 2625 GTATAGAGCATGGACAGACAGCGGACCAACACACACACACACAGGATGCCCGACATGCACA 2684
Db 2585 GTATAGGGCCCGG--GCCACCAATGTGTACACTGGGGCCACCAATCGTCCACC 2641
QY 2685 CCTGGATACAGAGAGGAACAATGACGACATTTATGTGTGGCTCCCGCCCCCAACACA 2744
Db 2642 CCAGTGTGCAG--CCTGAGGCTGGAGACTGGACCGCTGACTGCACCAAGCGCCC--CCAGA 2699
QY 2745 ACCGAGCTGTGAAGTCAATCTTAGGACTCAGAGT 2780
Db 2700 ACATACACTGTGAATCAATCTCCAGGCGCTCAAAAT 2735

RESULT 5
US-10-147-211A-1
; Sequence 1, Application US/10147211A
; Publication No. US20030235900A1
; GENERAL INFORMATION:
; APPLICANT: Yeh, Jiumn-Chern
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING A TRANSMEMBRANE SERINE PROTEASE 1
; TITLE OF INVENTION: ENCODED POLYPEPTIDES AND METHODS BASED THEREON
; FILE REFERENCE: 24745-1616
; CURRENT APPLICATION NUMBER: US/10/147,211A
; PRIOR FILING DATE: 2002-05-14
; PRIOR APPLICATION NUMBER: 60/291,001
; PRIOR FILING DATE: 2001-05-14
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: PastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 3147
; TYPE: DNA
; ORGANISM: Homo Sapien
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (23)...(2589)
; OTHER INFORMATION: Nucleotide sequence encoding MTSp1
; PUBLICATION INFORMATION:
; AUTHORS: O'Brien, T.J. and Tanimoto, H.
; DATABASE ACCESSION NUMBER: GenBank #AF081724
; DATABASE ENTRY DATE: 2000-08-31
; PATENT DOCUMENT NUMBER: 5,972,616
; PATENT FILING DATE: 1998-02-20
; PUBLICATION DATE: 1999-10-26
US-10-147-211A-1

Query Match 60.6%; Score 1883.2; DB 12; Length 3147;
Best Local Similarity 81.2%; Pred. No. 0;
Matches 2223; Conservative 0; Mismatches 508; Indels 5; Gaps 3;

QY 45 GATCGGACCGCCAAACCATGGTAGCAATCGGGCGCCGCAAGGCGGAGGGGGCTCTCAG 104
Db 5 GAGCGGCTCGGGGTACCATGGGAGCGATCGGGCCCGCAAGGGCGGAGGGGGCCGAG 64
QY 105 GACTTCGGCGGGGACTCAAGTACAACTCCCGGCTAGAGAACATGAATGGCTTTGAGAG 164
Db 65 GACTTCGGCGGGGACTCAAGTACAACTCCCGGACAGAGAAAGTGAATGGCTTGGAGGAA 124
QY 165 GGTGTGGAGTTCTCTGCTCGGACCAATCCCAAGAAAGTGGAGAGCGAGGCGCCAGCGC 224
Db 125 GCGGTGGAGTTCTCTGCCAGTCAACAACTCAAGAAAGTGGAGAAAGCATGGCCCGGGCGC 184
QY 225 TGGGTGGTGGTGGTGGAGTGTGTGAGTTCCTTGTCTCTCCCTCATGGCTGGCTTG 284
Db 185 TGGGTGGTGGTGGAGCGCTGTGATCGGCTCTCTTGGTCTTGGTGGGATCGGCTTC 244

QY 285 CTGGTGTGGCACTTCCATTATCGGAATGTGGGTTCAAAAAGTCTTCAATGGCCACTGT 344
Db 245 CTGGTGTGGCACTTTCAGTACCGGACGTGGTCCAGAGGCTTCAATGGCTACATG 304
QY 345 AGGATCAAAATGAGATCTTTCGTGATGCGTATGAGAACTCCACCTCCACAGATTTATC 404
Db 305 AGGATCAAAATGAGATTTTGTGGATGCTTACGAGAACTCCAACTCCACTGAGTTGTA 364
QY 405 AGCTGCCAGCAGGCTGAAGGAGGCGCTCAAGCTGTGTACAAATGAAGTCCCTGTCTGT 464
Db 365 AGCTGCCAGCAGGCTGAAGGAGGCGCTCAAGCTGTGTACAGCGAGTCCCAATTCCTG 424
QY 465 GGTCCCTTACCACAAGAGTGGCTGTAACTGCTTTCAGTGGAGGCGAGTGTATCGCCTAC 524
Db 425 GGCCCTTACCACAAGGAGTGGCTGTGACGCGCTTCAGCGAGGCGGCGTCAATCGCTAC 484
QY 525 TACTGTGAGATTCAGATCCCGCCACACTGCGCAGAGAGGTTGATCGGCGCATGGCT 584
Db 485 TACTGTGCTGAGTTTCAGCATCCCGCAGCACCTGTGGAGGAGCGCGCTCATGGCC 544
QY 585 GTGAGCGAGTTGTAACTTGCACCCCGCAGCAGCGGCACTGAAATCCTTTCGTCTAACA 644
Db 545 GAGGAGCGGTAGTATGCTGTGCCCCCGGGCGGCTCCCTGAGTCTCTTTGTGTCAAC 604
QY 645 TCTGTGTGGCTCTTCCCAATTGACCCCGAGAAATGCTGCAGAGGACTCAGGACAAACAGCTGC 704
Db 605 TCAGTGTGTGGCTTTCGCCACTCCAAAACAGTACAGAGGACCCAGGACAAACAGCTGC 664
QY 705 AGTTTTCCTGTGATGCCCATGTTGTCAGAGTGCACAGCTTCACTACCTTGGTTCCTCC 764
Db 665 AGCTTGGCTGTGACGCGCGGCTGTGGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGT 724
QY 765 AACAGTCCCTTACCCCGGCGCATGCCGTGTCAGTGGGCTCTGCGGGGGGAGCGCGACTCT 824
Db 725 GACAGCCCTTACCCCGCTCATGCCGTGTCAGTGGGCTCTGCGGGGGGAGCGCGACTCA 784
QY 825 GTGCTGAGCTCTACCTTCCGAAAGTTGATGTGCTCTCTGATGAGATGGCAGTGAC 884
Db 785 GTGCTGAGCTCTACCTTCCGAGCTTTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 844
QY 885 CTGGTCACTGCTGATGATGCTGAGCGGCTTGGAAACCCCAAGCTGCTGCTGCTGCTGCT 944
Db 845 CTGGTCACTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 904
QY 945 GGCACCTTCTCAGCTCTCCTTACAACTGACTTTCCTCTCTCTCTCTCTCTCTCTCTCTCT 1004
Db 905 GGCACCTTACCTTCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 964
QY 1005 AGCTGATTAACCAATACGACCGGCGACATCTGCTGCTTTCAGGCGCACTTCTTCTCAGCTG 1064
Db 965 ACATGATAACCAACATGAGCGCGGCTATCCCGGCTTTCAGGCGCACTTCTTCTCAGCTG 1024
QY 1065 CCAAGATGAGCAGCTGTGGCGGCTTTCAGTGCACACCCCAAGGACATTTAGCAGCCCC 1124
Db 1025 CTTAGGATGAGCAGCTGTGGAGCGGCTTACGTAAGCCCGAGGAGCACTTCAACAGCCCC 1084
QY 1125 TACTATCAGGCGCACTACCCCGCCCAACATCACTCACTGCAATATCAAGAGTGGCCCAAC 1184
Db 1085 TACTATCAGGCGCACTACCCCAACATCACTGCAATATCAAGAGTGGCCCAAC 1144
QY 1185 AACGGAACTGAGGTCGCTTCAAACTCTTCTATCTGCTGAGACCCCAAGTACCAAGT 1244
Db 1145 AACGAGATGAGGTCGCTTCAAACTCTTCTATCTGCTGAGCGCGGCTGCTGCTGCTGCTGCT 1204
QY 1245 GGCTCTGTCACCAAGGACTATGTGGAGATCAACGGGAGAGTACTCGGCTGAGAGTCC 1304
Db 1205 GGCACTTCCCGCAAGGACTAGTGGAGATCAATGGGAGAAATACTCGGAGAGAGTCC 1264
QY 1305 CAGTTTGTGGTGGAGCAACAGCAGCAAGATTACAGTCCCACTTCCATTTCTGATCACTCG 1364
Db 1265 CAGTTTGTGGTGGAGCAACAGCAGCAAGATTACAGTTCGCTTCCACTCAGATCAGTCC 1324

QY 1365 TACACGACACCGGTTCTTAGTACGACCTCTCTAGTACCTCAACAGCCCGTGCACA 1424
DB 1325 TACACGACACCGGTTCTTAGTACCTCAACCTCTCTAGTACCTCAACAGCCCGTGC 1384
QY 1425 GGGATGTTATGTCAGACACTGAGCGGTGCATCCGAAAGGAACCTGCGCTGCGACGGCTGG 1484
DB 1385 GGGCGATTACGTGCGCACCGGGGGGGTGTATCCGGAAGGAGCTGCGCTGTGATGGCTGG 1444
QY 1485 GCAGACTCCCGGATTTATAGTATGATGAGCGGTACTGCGGATGCAATGCCACCCAGCTTC 1544
DB 1445 GCCGACTGACCGACACAGCGATGAGCTCACTGAGTTGGAGCGCGCCACCGATTC 1504
QY 1545 ACGTCAAAAACCAAGTTCTGCAAGCCCTCTTCTCGGTTCTGTGACAGTGTCAACAGCTGT 1604
DB 1505 ACGTCAAGAACCAAGTTCTGCAAGCCCTCTTCTCGGTTCTGTGACAGTGTCAACAGCTGT 1564
QY 1605 GGGAGCGAAGTGTGAGGAGGGCTGACGTGCTCTGCTGGAGTTTCAAGTGTCCAAAT 1664
DB 1565 GAGACAAACAGCGACGAGCAGGGGTGCAGTTGTCCGGGCCAGACCTTCAGGTTGTTCCAAAT 1624
QY 1665 GGGAAAGTGTCTCCCTCAGAGCCAGAAAGTGTATGGAAAGGACAACTGTGGAGATGGGTCT 1724
DB 1625 GGGAAAGTGTCTCCGAAAGCCAGCAGTGCATGGGAAGGAGCACTGTGGGACCGGTTCC 1684
QY 1725 GACGAGGTTCAATGTGACAGCGTGAATGTCGTCTTGTGACCAAAATATACCTACCGCTGC 1784
DB 1685 GACGAGGCTCTGTCGCCCAAGGTGAAGCTGCTGCTTGTATACCAACACACCTTACCGCTGC 1744
QY 1785 CAAAATGGCTCTGCTGAGCAAGGCAACCTGAGTGTGATGGGAAGCAGACGTGTAGC 1844
DB 1745 CTCATGGGCTCTGTTGAGCAAGGCAACCTGAGTGTGAGGGAAGGAGACTGTAGC 1804
QY 1845 GATGCTCCGATGAGAAAACCTGTGACTGTGGGCTGCGATCTCTTACCAAAACAGAGCTGCG 1904
DB 1805 GACGCTCAGATGAGAGCACTGCGACTGTGGGCTGCGGTCAATTCAGGACAGAGCTGT 1864
QY 1905 GTGGTTGTGGCAAGAAATGCGAGCAGGCGAGTGGCCCTGCGAGGTGAGCTTCACGCGC 1964
DB 1865 GTTGTGGGGGACGATGCGGATGAGGCGAGTGGCCCTGCGAGGTAAAGCTGATGCT 1924
QY 1965 CTGGCCAGGGCCACTGTGTGGGCTCGCTCATCTCTCTGACTGCTGCTGCTCTGCA 2024
DB 1925 CTGGCCAGGGCCACTGTGTGGGCTCGCTCATCTCTCTGACTGCTGCTGCTCTGCTG 1984
QY 2025 GTCATGCTTTCAGGATGACAAAAATTTCAAGTACTCAGACTACACGATGTGGACGGCC 2084
DB 1985 GCACACTGCTACATGATGACAGAGATTCAGTACTCAGACCCACGAGTGGAGCGCC 2044
QY 2085 TTCCTGGGTCTGCTGGACAGAGCAAGCGCAGTGTCTTGGGGTGAGGAGCTGAAGCTC 2144
DB 2045 TTCCTGGGCTTGCAGACCCAGAGCCAGCGCGCCCTTGGGGTGAGGAGCGCAGGCTC 2104
QY 2145 AAAAGTATCATCACCCCTCTTCTCAATGATTTCACTTCGACTATGATGATGATGATGAT 2204
DB 2105 AAGCGATCATCTCCACCCCTCTTCTCAATGATTTCACTTCGACTATGATGATGATGAT 2164
QY 2205 CTGGAGCTGGAGAAAGTTCGGTGAAGTACAGCAGCGTGTGCGGCCCATCTGCTGCTGCTGAT 2264
DB 2165 CTGGAGCTGGAGAAACCGCAGATACAGTTCATGCTGCGGCCCATCTGCTGCTGCGGAC 2224
QY 2265 GCTACCATGTCTTCTGCTGCAAGCGCATCTGCGGTCAAGGCTGGGGGCAACAAAA 2324
DB 2225 GCTTCCCATGTCTTCTGCTGCAAGCGCATCTGCGGTCAAGGCTGGGGGCAACACCCAG 2284
QY 2325 GAGGAGGTACCGGAGCGCTGATCTCTGAGAGGGTGTGATCCGTGTATCAACAGGAC 2384
DB 2285 TATGAGGACATGCGGCGCTGATCTCTGCAAAAGGGTGTGATCCGTGTATCAACAGGAC 2344
QY 2385 ACTGTGAGGACCTCATGCGCGCAGAGATCAACCCACGAATGATGTGTGGGTTCTCTC 2444
DB 2345 ACTGCGAGAACTCTCTGCGGACGAGATCAACCCCGGATGATGTGTGGGTTCTCTC 2404
QY 2445 AGTGGGGGTGTGGAATCTCTGCGCAGGGGTGACTCTGGTGGCCCCCTTGTCAAGCGCGGAGAA 2504

DB 2405 AGCGCGCGGTGGACTCTGTCAGGGTATTCCGGGGGACCCCTCTCTCAGCGTGGAGGG 2464
QY 2505 GATGGGCGAATGTTCCAGGCTGTGTGTGAGCTGGGTGAAGGTGCGCTCAGAGGAAC 2564
DB 2465 GATGGGCGGATCTTCCAGCGCGGTGTGTGTGAGCTGGGGAGACGGCTGCGCTCAGAGGAAC 2524
QY 2565 AAGCCAGCGGTGTACAAAGGCTCCCTGTAGTTGGGACTGATCAAAAGACACACTGG 2624
DB 2525 AAGCCAGCGGTGTACAAAGGCTCCCTGTAGTTGGGACTGATCAAAAGACACACTGG 2584
QY 2625 GTATAGCAGTGTGACAGACAGCCGACCAACAAACACCCACAGGGATGCCGACATGCACA 2684
DB 2585 GTATAGGGGCGGG---GCCACCCAAATGTGTACACTGCGGGGCCACCCATCGTCCACC 2641
QY 2685 CCTGATACAGAGAGGAGAACTGACGACATTTATGCTGTGCTCCCTCCCTCCCAACACA 2744
DB 2642 CAGTGTGCAGC-CCTGAGGCTGGAGACTGGACGCTGACTGACACAGCGGCC-CCAGA 2699
QY 2745 ACCCAGACTGTGAATGCTGATCTTTAGGACTCAGAGT 2780
DB 2700 ACATACACTGTGAATCAATCTCCAGGGCTCCAAAT 2735

RESULT 6

US-10-147-211A-3
; Sequence 3, Application US/10147211A
; Publication No. US20030235900A1
; GENERAL INFORMATION:
; APPLICANT: Madison, Edward
; APPLICANT: Yeh, Jium-Chern
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING A TRANSMEMBRANE SERINE PROTEASE 1:
; TITLE OF INVENTION: ENCODED POLYPEPTIDES AND METHODS BASED THEREON
; FILE REFERENCE: 24745-1616
; CURRENT APPLICATION NUMBER: US/10/147,211A
; CURRENT FILING DATE: 2002-05-14
; PRIOR APPLICATION NUMBER: 60/291,001
; PRIOR FILING DATE: 2001-05-14
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 3147
; TYPE: DNA
; ORGANISM: Homo Sapien
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1865)...(2590)
; OTHER INFORMATION: Nucleic acid sequence of protease domain of MTSP1
US-10-147-211A-3

Query Match 60.6%; Score 1883.2; DB 12; Length 3147;
Best Local Similarity 81.2%; Pred. No. 0;
Matches 2223; Conservative 0; Mismatches 508; Indels 5; Gaps 3;
QY 45 GATCGACCGCCAAACCATGGGTAGCAATCGGGCCCGCAAGCGCGAGGGGGCTCTCAG 104
DB 5 GAGCGGCTTCGGGGTACCATGGGAGCGATCGGGCCCGCAAGGGCGAGGGGGCCGAAG 64
QY 105 GACTTCGGCGGGGACTCAAGTACAACTCCCGGTAGGACATGAATGGCTTTGAGGAG 164
DB 65 GACTTCGGCGGGGACTCAAGTACAACTCCCGGACGAGAAAGTGAATGGCTTTGAGGAA 124
QY 165 GGTGTGAGTTCCTGCTCGCAACAAATGCCAAGAAAGTGGAGAGCGAGGCCCGCCAGGCGC 224
DB 125 GCGGTGGAGTTCCTGCCAGTCAACAACTCAAGAGGTGGAAAGCATGGCCCGGGGGCGC 184
QY 225 TGGGTGCTGCTGGTGGCAGTGTGTTCAGTTCCTCTTGTCTCTCTCTCTCTCTCTCTCT 284
DB 185 TGGGTGCTGCTGGCAGCGCTGTGTATCGGCTCTCTTGTCTCTCTCTCTCTCTCTCTCT 244
QY 285 CTGGTGTGGCACTTCCATATCGGNAATGTGGGGTTCAAAAGTCTTCAATGGCCATCTG 344
DB 245 CTGGTGTGGCACTTTCAGTACCGGAGCGTGTGTCCAGAAAGGTCTTCAATGGCTACATG 304

QY 345 AGGATCAAAATGAGATCTTTCTGGATCGGTATGAGAACTCCACTCCACAGAGTTTATC 404
Db 305 AGGATCAAAATGAGAAATTTTGTGGATGCTTACGAGAACTCCAATCCACTGAGTTTGTGA 364
QY 405 AGCTGGCCAGCGAGGTGAAGGAGCGGTGAAGCTGTGTAAATGAAGTCCCTGTCTGTG 464
Db 365 AGCTGGCCAGCGAGGTGAAGGAGCGGTGAAGCTGTGTAAATGAAGTCCCTGTCTGTG 424
QY 465 GGTCCCTACCAACAAGAGTCCGCTGTAACTGCTTCACTGAGGGGAGGTGATCGCCCTAC 524
Db 425 GGCCCTACCAACAAGAGTCCGCTGTAACTGCTTCACTGAGGGGAGGTGATCGCCCTAC 484
QY 525 TACTGGTCAGAGTTTCAAGTATCCCCCAACCTTGGCAGAGAGGTGTGATCGGCCCATGGCT 584
Db 485 TACTGGTCAGAGTTTCAAGTATCCCCCAACCTTGGCAGAGAGGTGATCGGCCCATGGCT 544
QY 585 GTGGAGCAGTTGTAACTTGGCAGCCCGAGCAGCGGCACTGAAATCTTCTGTGTAAACA 644
Db 545 GAGGAGCGGTAGTATGCTGCCCCCGGGCGGCTCCCTGAAATCTTCTGTGTAAACA 604
QY 645 TCTGTGTGGCTTCCCAATGACCCCAAGATGCTGACAGAGCTTCAGGACAAACAGCTGC 704
Db 605 TCAGTTGGTGGCTTTCCCAAGACTTCCAAACAGTACAGAGACCCAGGACAAACAGCTGC 664
QY 705 AGTTTGGCTTGCATGCCCCATGGTGCAGCAGTGAACAGCTTCACTTACCCCTGGCTTCCCC 764
Db 665 AGCTTGGCTTGCATGCCCCCGGGTGTGAGCTGATGCGCTTCAACACGCCCGCTTCCCT 724
QY 765 AACAGTCTTACCCCGCGATGCCCCGTGCGAGTGGTCTGCGGGGGAGCGCGACTCT 824
Db 725 GACAGCCCTTACCCCGCTGATGCCCCGTGCGAGTGGGCTGCGGGGGAGCGCGACTCA 784
QY 825 GTGTGAGCTTCACTTCCGAAGCTTTCATGTCGCTTCCCTGTGATGAGTGCAGTGCAC 884
Db 785 GTGTGAGCTTCACTTCCGAAGCTTTCATGTCGCTTCCCTGTGATGAGTGCAGTGCAC 844
QY 885 CTGGTCAACCGTGTATGATAGCTGAGCCCCCATGGAACCCCAAGCTGTGTGGGGTGTGT 944
Db 845 CTGGTCAACCGTGTATGATAGCTGAGCCCCCATGGAACCCCAAGCTGTGTGGGGTGTGT 904
QY 945 GGCACTTCTTCACTTCTTCACTGAGTTCCTTCTCTCCAGAAAGTCTTCTTCTGTC 1004
Db 905 GGCACTTCTTCACTTCTTCACTGAGTTCCTTCTCTCCAGAAAGTCTTCTTCTGTC 964
QY 1005 ACCTGTATAACCAATACCTGACCGCGCATCTCTGCTTTGAGGCCACTTCTTCCAGCTG 1064
Db 965 ACCTGTATAACCAATACCTGAGCGCGCGCATCCCGCTTTGAGGCCACTTCTTCCAGCTG 1024
QY 1065 CCCAAGATGAGAGTGTGGCGGCTTTTGTAGTGACA CCCAAGGACATTTAGCAGCCCC 1124
Db 1025 CCTAGGATGAGAGTGTGGAGCGCTTACGTAAGGCCCGAGGACATTTCAACAGCCCC 1084
QY 1125 TACTATCAGGCCACTACCCGCCCAACATCACTGACATGAAATATCAAGTGCCTCAAC 1184
Db 1085 TACTATCAGGCCACTACCCGCCCAACATGACTGACATGAAATATCAAGTGCCTCAAC 1144
QY 1185 AACCGGAACGTGAAGGTGGCTTCAAACTTCTTATCTGTGTGAGCCCCCAAGTACAGTG 1244
Db 1145 AACAGCATGTGAAGGTGAGCTTCAAAATTTCTTACCTGCTGGAGCCCGGGTGCCTGG 1204
QY 1245 GGCTCTGCAACAGACTATGTGAGATCAACGGGGAGAGTACTGGGGTGAAGGTTC 1304
Db 1205 GGCACTTGGCCCCAAGACTTACCTGAGATCAATGGGGAAATACTGGGAGAGAGGTTC 1264
QY 1305 CAGTTTGTGTGAGCAGCAACAGCAGCAAGATTACAGTCCACTTCCATTTCTGATCACTCG 1364
Db 1265 CAGTTTGTGTGATCACCAGCAACAGCAGCAAGATCAAGTTCCTTCCACTCAGTCACTCC 1324
QY 1365 TACACGGACACCGGGTTCCTAGCTGAGTACCTTCTCTAGCACTCAACAGACCCCGGCCCA 1424
Db 1325 TACACGGACACCGGGTTCCTTACCTGAATACCTTCTCTAGCACTCCAGTGAACCATGCCGG 1384

QY 1425 GGGATGTTCTATGTGAAAGTGGACGGTGCATCCGAAGGAACTGCGCTGCGACGGCTGG 1484
Db 1385 GGGCAGTTTCACTGTCGCCACCGGGCGGTGTATCCGAAGGAGCTGCGCTGTGATGGCTGG 1444
QY 1485 GCAGACTGCCCGGATTTATGTATGAGCGTTTATCTGCGATGCAATGCCACCCACAGTTTC 1544
Db 1445 GCCAGCTGCACCGACCAACAGCGATGAGCTCAATGTCAGTTTGCAGCGCGCGCCACAGTTTC 1504
QY 1545 ACCTGCAAAACACAGTTTCTCAAGCCCTCTTCTGGGTCTGTGACAGTGTCAACGACTGT 1604
Db 1505 ACCTGCAAAACACAGTTTCTCAAGCCCTCTTCTGGGTCTGTGACAGTGTGAACGACTGC 1564
QY 1605 GGGACAGAAAGTACGAGGAGGGCTGCAGCTGTCTCTGCTGGGAGTTTCAAGTGTTCAT 1664
Db 1565 GGAGCAACACGACGAGCAGGGGTGCAGTTGTCCGGCCAGACCTTCAGGTGTTCAT 1624
QY 1665 GGGAGGTCTTCTCTCCTCAGACGAGAGTGTAAATGGGAAGCACTGTGTGAGATGGGTCT 1724
Db 1625 GGGAGGTCTTCTCTCCTCAGAAACCCAGCAGTGCATGGAAGGAGCGACTGTGGGGACGGGTTC 1684
QY 1725 GACGAGCTTCACTGACAGCGTGAATGTCTCTTGTGACCAAAATATACCTACCGCTGC 1784
Db 1685 GACGAGCTTCTCTCCTCAGCGGAGTGAAGTGTCACTTGTACCAAAACACACTACCGCTGC 1744
QY 1785 CAAAATGGCTTCTCTGTAGCAAGGCAACCTCAGTGTGTATGGGAAGAGAGCTGTAGC 1844
Db 1745 CTCAATGGCTTCTCTGTAGCAAGGCAACCTCAGTGTGTAGCGGAGAGGAGCTGTAGC 1804
QY 1845 GATGGCTCCGATGAGAAACTGTGATGTGGGCTGCGATCTTTTACCMACAGCTGC 1904
Db 1805 GACGGCTCAGATGAGAAAGTGTGACTGTGGGCTGGGTCAATTCAGAGACAGGCTGT 1864
QY 1905 GTGGTGTGTGCAACGAATGCGGACGAGGGGAGTGGCCCTGGCAGGTGAGCTTCCACGCCC 1964
Db 1865 GTGGTGTGTGCAACGAATGCGGACGAGGGGAGTGGCCCTGGCAGGTGAGCTTCCACGCCC 1924
QY 1965 CTGGGCGAGGCCACTTGTGTGGGCTGTCTCATCTCTCCAACTGGCTGTCTGCA 2024
Db 1925 CTGGGCGAGGCCACTTGTGGTGTCTTCCCTCATCTCTCCAACTGGCTGTCTGCA 1984
QY 2025 GCTCATTTGCTTTCAGGATGACAAAATTTCAAGTACTCAGACTACAGATGTGAGCGGCC 2084
Db 1985 GCACACTGCTACATCGATGACAGAGGATTCAGGTACTCAGACCCCGCAGTGCAGCGGCC 2044
QY 2085 TTCTGTGTGTGTGCAACGAGCAAGCGAGTGCCTTCTGGGGTGCAGGAGCTTCAAGCTC 2144
Db 2045 TTCTGTGTGTGTGCAACGAGCGAGCGAGCGCCCTGGGGTGCAGGAGCGCGAGCTC 2104
QY 2145 AAAAGTATCATCACCCACTTCTTCAATGATTTTCACTTGTGACTATGACATCGCTTG 2204
Db 2105 AAGGCGATCATCTCCACCCCTTCTTCAATGACTTCACTTTCGACTATGACATCGCGCTG 2164
QY 2205 CTGGAGCTGGAGAGTTCGGTGTGAGTACAGACCGCTGTGGGCCCCCTCTGCTGCTGAT 2264
Db 2165 CTGGAGCTGGAGAGTTCGGTGTGAGTACAGTTCATGTGCGGCCCTCTGCTGCGGAC 2224
QY 2265 GCTACCCATGCTTCTTCTGTGCAAGGCCATCTGGGTTCAGAGCTGGGGGCAACAAAA 2324
Db 2225 GCTTCCATGCTTCTTCTGCGGCAAGGCCATCTGGGTTCAGAGCTGGGGGCAACAAAA 2284
QY 2325 GAGGAGGTACCGGAGCGCTGATCTGCGAGAGGGTGTGATCCGTGTCTCAACAGAGCC 2384
Db 2285 TATGAGGCACTGCGCGCTGATCTTCAAAAGGGTGTGATCCCGCTCATCAACAGAGCC 2344
QY 2385 ACCTGTGAGGACCTCATGCGCGCAGAGATCAACCCCAAGATGATGTGTGGGTTCCTC 2444
Db 2345 ACCTGTGAGGACCTCATGCGCGCAGAGATCAACCCCGCGCATGTGTGGGTTCCTC 2404
QY 2445 AGTGGGGTGTGAGTCTCTGCGAGGGTGTCTGTGGTGGCCCTTGTCAAGCGCGGAGAA 2504
Db 2405 AGCGGCGGTGGACTCTCTGCGAGGGTGTATTCGCGGGGACCCCTGTCCAGCGTGGAGCG 2464
QY 2505 GATGGCGAATGTTCCAGGGCTGTGTGTGAGCTGGGGTGAAGGCTCGGCTCAGAGNAC 2564

Db	2465	GATGGGCGGATCTTCAGGCCGTGTGGTGGAGCGGCTCGCTCAGGGAAC	2524
QY	2565	AAGCCAGGCGTGATACACAAGGCTCCCTGTAGTTTCGGGACTTGGATCAAGAGCACATGGG	2624
Db	2525	AAGCCAGGCGGTGTACACAAGGCTCCCTCTGTTTGGGACTTGGATCAAGAGAGACATCTGG	2584
QY	2625	GTATAGCAGCATGTGACAGACAGCCGACCAAAACACCACAGGAGTGCCTCGAATGCACA	2684
Db	2585	GTATAGGGCGGG--GCCACCAAATGTGTACCTTCGGGGCCACCCTCGTCCACC	2641
QY	2685	CCTTGGATACAGGAGAGAACTGACGACATTTATGTGTGGCTCCCCCCCCCAACACA	2744
Db	2642	CCAGTGTGCAG--CCTCAGGCTGGAGCTGGACCGCTGACTGCACAGCGCCC--CCAGA	2699
QY	2745	ACCAGACTGTGAATGCATCCTTTAGGACTCAGAGT	2780
Db	2700	ACATACACTGTGAATCAATCTCCAGGGCTCCAAAT	2735

```

RESULT 7
US-10-156-214A-1
; Sequence 1, Application US/10156214A
; Publication No. US2004001801A1
; GENERAL INFORMATION:
; APPLICANT: Edwin L. Madison
; APPLICANT: Joseph Edward Semple
; APPLICANT: George P. Vlasuk
; APPLICANT: Scott Jeffrey Kemp
; APPLICANT: Mallareddy Komandla
; APPLICANT: Daniel Vanna Siev
; TITLE OF INVENTION: Conjugates Activated By Cell Surface Proteases and Therapeutic Us
; TITLE OF INVENTION: Thereof
; FILE REFERENCE: 24745-1611
; CURRENT APPLICATION NUMBER: US/10/156,214A
; CURRENT FILING DATE: 2002-05-23
; NUMBER OF SEQ ID NOS: 611
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 3147
; TYPE: DNA
; ORGANISM: Homo Sapien
; FEATURE:
; NAME/KEY: CDS
; LOCATION: {23}...(2589)
; OTHER INFORMATION: Nucleotide sequence encoding MTSPI
; PUBLICATION INFORMATION:
; AUTHORS: O'Brien, T. J. and Tanimoto, H.
; DATABASE ACCESSION NUMBER: GenBank AR081724
; DATABASE ENTRY DATE: 2000-08-31
; PATENT DOCUMENT NUMBER: US Pat 5972616
; PATENT FILING DATE: 1998-02-20
; PUBLICATION DATE: 1999-10-26
; US-10-156-214A-1

```

185	Db	TGGGTGGTCTGGCAGCCGTTGCTGATCGGCCTCCTCTTGGTCTTGGTATCGGGATCGGGCTTC	244
285	QY	CTGGTGTGCACCTTCCATATATCGGAATGTGCGGGTTCAAAAAGTCTTTCAATGCCCATCTG	344
245	Db	CTGGTGTGGCATTTGTCAGTACCGGGACGTGCGGTGCCAGAGGTCTTCAATGGCTACATG	304
345	QY	AGGATCACAAATCAGATCTTTCTGGATGGTATGAACTCCACACTCCACAGAGTTTATC	404
305	Db	AGGATCACAAATGAGAAATTTTGTGGATGCTACGAGAACTCCAACTCCCATGTAGTTTGA	364
405	QY	AGCCTCGCCAGCCAGGTGAAGGAGGCGCTGAAGCTCTGTCAAAATGAAGTCCCTGTCTCCTG	464
365	Db	AGCCTCGCCAGCAGGTGAAGGACGCGCTGAAGCTCTGTACAGCGAGTCCCATTTCCCTG	424
465	QY	GGTCCCTACACAAGAAGTCGGCTGTAACTGCTTCAGTGAGGGCAGTGTCACTCGGCTTAC	524
425	Db	GGCCCCCTACACAAGAGATCGGCTGTGTGAGCGCTTTCAGCGAGGGCAGCTCATCGGCTAC	484
525	QY	TACTGTGTCAAGTTTCAGCATCCCCCACCTCGGCAGAGAAGGTTTCATCGCGCATGGCT	584
485	Db	TACTGTGTGTAGTTTACGATCCCGCAGACCTTGGTGGAGAGGCCGAGCGGTCACTGGCC	544
585	QY	GTGGAGCGAGTTGTAACTATGTCACCCCGAGCAGCGGCACTGAAATCTTTGTTGTAAACA	644
545	Db	GAGGAGCGGTAGTCTCATGTGCCCCCGCGGGCGGCTCCCTGAAAGTCTTTGTGTGTCACC	604
645	QY	TCTGTGCTGGCTTCCCCCATTTGACCCCCAGAACTGTCAGAGAGACTCAGAGACAAACAGCTGC	704
605	Db	TCAGTGGTGGCTTTCCCCAGGACTCAAAACAGTACAGAGAGCCAGGACAAACAGCTGC	664
705	QY	AGTTTTGGCCTGCATGCCCATGGTGCAGCAGTGAACGCTTTCACACTACCCCTGGCTTCCCC	764
665	Db	AGCTTTGGCTGTCACGCCCGCGTGTGGAGCTGATGCGCTTTCACCAACGCCGCGTTCCTT	724
765	QY	AACAGTCCCTACCCGGGGCATGCCGCTGGCAGTGGTCTCTCGGGGGGACGCCGACTCT	824
725	Db	GACAGCCCCCTACCCCGCTCATGCCCGCTGCGAGTGGGCCCCCTGCGGGGGGACGCCGACTCA	784
825	QY	GTGCTGAGCCTCACCTTCCGAAGCTTTGATGTGCTTCCCTCTGTGATGAGCATGGCAGTGAC	884
785	Db	GTGCTGAGCCTCACCTTCCGAGCTTTTGACCTTGGCTCTGCGACGAGCGCGGACGCGAC	844
885	QY	CTGGTCACCGTGTATGATAGCTGAGCCCCCATGGAACCCCAACGCTGTGGTGGCGGTGTGT	944
845	Db	CTGGTGAACGCTGTACAAACACCTGAGCCCCCATGGAGCCCCCAACGCCCCCTGGTGCAGTTGTGT	904
945	QY	GGCAGCTTCTCACCCCTCTACAACTGATCTTCTCTCTCTCCAGAAAGTCTTCTCTGTCTC	1004
905	Db	GGCACCTTACCTTCCCTTCTACAACTGACTTTCCTTCCCTCCAGAAACGCTCTGCTCATC	964
1005	QY	AGCCTGATAACCAATACTGACCGGGACATCTCTGGCTTTGAGGCCACTTTCTTCTCCAGCTG	1064
965	Db	ACACTGATAACCAACACTGAGCGGGCGCATCCCGCTTTGAGGCCCACTTCTTCTCCAGCTG	1024
1065	QY	CCCAAGATGAGCAGCTGTGCGGCTTTTGTGAGTGACACCAAGGGACATTTAGCAGCCCC	1124
1025	Db	CCTAGGATGAGCAGCTGTGAGGCGCCTTACGTAAAGCCGAGGGACATTCACAGCCCC	1084
1125	QY	TACTATCCAGGCCACTACCCGCCCAACATCAACTGACATGGAATATCAAGGTGCCCAAC	1184
1085	Db	TACTATCCAGGCCACTACCCACCCAACTTGAAGCTGACATGGAACATTCAGGTGCCCAAC	1144
1185	QY	AACCGAAAGTGAAGTGGCTTCAAACTCTTCTATCTGGTGGACCCCAACGTTACAGTG	1244
1145	Db	AACGACATGTGAAGTGAAGTTCAAAATTTCTTCTTCTCTCTGAGGCCGCGGCTGGCG	1204
1245	QY	GGCTCTCTGACCAAGGACTATGTGAGAGTCAACCGGGGAAAGTACTTCGGTGAAGGCTCC	1304
1205	Db	GGCACCTGCCCCAGGACTACTGTTGGAGTCAATGGGAGAAATACTTCGGAGNAGAGTCC	1264
1305	QY	CAGTTTGTGGTGAAGCAGCAACAGCAGCAAGATTTACAGTCCACTTCCATTTCTGATCACTCG	1364
1265	Db	CAGTTTCGTCTACCAAGCAACAGCAACAGATCAAGTTTCGTTCCCTTCCACTCAGATCAGTCC	1324


```
QY 1365 TACACGACACCGGTTCTTACGTGAGTACCTTCTCTACGACTCCAAAGACCCCGTCCCA 1424
Db 1325 TACACGACACCGGTTCTTACGTGAGTACCTTCTCTACGACTCCAAAGACCCCGTCCCG 1384
QY 1425 GGGATGTTTCATGTGCAAGACTGGACGCTGCATCCGAAAGAACTGCGTGGACGCGTGG 1484
Db 1395 GGGAGTTTCACTGTCGCGACGCGGCGGTGTATCCGAAAGAACTGCGTGGTGTGATGGTGG 1444
QY 1485 GCAGACTCCCGGATATAGTGATGAGGGTTACTGCCGATGCAATGCCACCCACCACTTC 1544
Db 1445 GCGGACTGCACCGACACACAGGATGAGCTCAACTGCAGTTGGACGCGGCCACCACTTC 1504
QY 1545 ACCTGCAAAACCACTTCTGACAGCCCTTCTTGGGTCTGTGACAGTGTCAAACAGTGT 1604
Db 1505 ACCTGCAAAACCACTTCTGACAGCCCTTCTTGGGTCTGTGACAGTGTCAAACAGTGT 1564
QY 1605 GGGGACGGAAGTGACAGAGGGCTGCAGCTGCTCTGCTGGGAGTTTCAAGTGTTCGAAT 1664
Db 1565 GGAGACAAACAGCAGCAGGAGTGTGAGTGTGCTGGGCCAGACCTTCAGTGTTCGAAT 1624
QY 1665 GGGAGTGTCTCCCTCAGAGCCAGAAAGTGTAAATGGAAAGCAACTGTGGAGATGGTCT 1724
Db 1625 GGGAGTGTCTCTCGAAAGCAGCAGTGTCAATGGAAAGCAACTGTGGGAGCGGTCC 1684
QY 1725 GACGAGGCTTCATGTGACAGCGTGAATGTCGTCTCTGCTGGGAGTTTCAAGTGTTCGAAT 1784
Db 1685 GACGAGGCTTCCTGCCCAAGGTGAACGTGCTCACTTGTACCAACACACCTTACCGCTGC 1744
QY 1785 CAAATATGCTCTGTCTGACGAAAGGCAACCTCGAGTGTGATGGAAAGCAAGCACTGTAGC 1844
Db 1745 CTCATATGGCTCTGCTTGAAGAGGCAACCTCGAGTGTGATGGAAAGCAAGCACTGTAGC 1804
QY 1845 GATGCTCCGATGAGAAACCTGACCTGTGGGTGCGATCTTTTACCAACAGGCTCGC 1904
Db 1805 GACGCTCAGATGAGAAAGCACTGCGACTGTGGGTGCGCTCATTCACGAGACAGGCTCGT 1864
QY 1905 GTGGTTGTGGCAAGTACGACAGCGGAGTGGCGCTGCGCTGCGAGTGTGAGCTTCCAGCC 1964
Db 1865 GTTGTGGGGCAGGATGCGGATGAGGCGAGTGGCGCTGCGCTGCGAGTGTGAGCTTCCAGCT 1924
QY 1965 CTGGCCACGAGGCACTGTGTGGGCGCTCGCTCATCTCTCTGACTGTGCTGTCTCTCA 2024
Db 1925 CTGGCCACGAGGCACTGTGGGTGCTCTCTCTCATCTCTCCAACTGCTGTGCTCTGCC 1984
QY 2025 GCTCATCTCTTACGATGACAAAATTTCAAGTACTCAGACTACACGATGTGACGCGCC 2084
Db 1985 GCACACTGCTATCATCGATGACAGAGGATTCAGTACTCAGACCCCAAGTGTGACGCGCC 2044
QY 2085 TTCTGGGTCTGCTGACGACAGCAGCGAGTGGCTCTGGGTGCGAGGCTGAGGCTC 2144
Db 2045 TTCTGGGTCTGCGACGACAGCGAGCGAGCGCGCTGCGGTGCGAGGCTGAGGCTC 2104
QY 2145 AAACGTATCATCACCAACCTTCTTCAATGATTTTCACTTTCGACTATGATCATCGCTTG 2204
Db 2105 AAGCGCATCATCTCCACCCCTTCTTCAATGATTTTCACTTTCGACTATGATCATCGCTTG 2164
QY 2205 CTGGAGCTGGAGAACTGGTGGAGTACAGACCGTGTGTCGCGCCCATCTGCTGCTGAT 2264
Db 2165 CTGGAGCTGGAGAAACCGGACAGTACAGCTCCATGGTGGCGCCCATCTGCTGCGGAC 2224
QY 2265 GCTACCCATGCTTCCCTGCTGGCAGGCGCATCTGGTTCAGAGCTGGGGGACACAAA 2324
Db 2225 GCTTCCCATGCTTCCCTGCTGGCAGGCGCATCTGGTTCAGAGCTGGGGGACACACCAG 2284
QY 2325 GAGGAGGTACCGGAGCGCTGATCTCTGCAAGAGGTGAGATCCGCTGTCTATCAACACAGACC 2384
Db 2285 TATGAGGCACTGGCGCTGATCTGCAAGAGGTGAGATCCGCTGTCTATCAACACAGACC 2344
QY 2385 ACCTGTGAGGACCTCATGCGCAGCAGATCACCCCAAGATGATGTGTGGGTTCCTC 2444
Db 2345 ACCTGTGAGAACCTCTGCGCAGCAGATCACCGCGCATGATGTGTGGGTTCCTC 2404
```

```
QY 2445 AGTGGGTTGTGACTCTCTGCCAGGTGACTCTGTGTGCGCCCTTGTCAAGCGCGAGAAA 2504
Db 2405 AGCGGCGCGTGGACTCTCTGCCAGGTGACTCTGTGTGCGGAGACCCCTGTCCAGCGTGGAGCG 2464
QY 2505 GATGGCGCAATGTTTCCAGGCTGTGTGTGAGCTGGGTGAGGCTGCGCTCAGAGGAAC 2564
Db 2465 GATGGCGGATCTTTCAGGCGCGGTGTGTGAGCTGGGAGACGCTGCGCTCAGAGGAAC 2524
QY 2565 AAGCAGCGGTGTACAAAGGCTCCCTGTGTGTGAGTGTGATCAAAAGACCACTGGG 2624
Db 2525 AAGCAGCGGTGTACAAAGGCTCCCTGTGTGTGAGTGTGATCAAAAGACCACTGGG 2584
QY 2625 GTATAGCAGTGTGACAGACAGCGACCAAAACACCCACAGGATGCCACATGCA 2684
Db 2585 GTATAGGCGCGCGG---GCCACCCAAATGTGTACCTGTGCGGCGCACCCATCGTCCACC 2641
QY 2685 CCTGATACAGGAGAGGAAACACTGACGACATTTATGTGTGCTTCCCGCCCAACACA 2744
Db 2642 CCAGTGTGACG-CCTGACGCTGGAGCTGGACCGCTGACTGCACACAGCGCCC-CCAGA 2699
QY 2745 ACCCAGACTGTGAATGCATCTTTAGGACTCAGAGT 2780
Db 2700 ACATACACTGTGAATCAATCTCCAGGCGCTCCAAAT 2735
```

RESULT 8

```
US-10-190-030B-1
; Sequence 1, Application US/10190030B
; Publication No. US20030134298A1
; GENERAL INFORMATION:
; APPLICANT: Madison, Edwin
; APPLICANT: Ong, Edgar
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING A
; TITLE OF INVENTION: TRANSMEMBRANE SERINE PROTEASE 20, THE ENCODED POLYPEPTIDES AND
; FILE REFERENCE: 24745-1618
; CURRENT APPLICATION NUMBER: US/10/190,030B
; CURRENT FILING DATE: 2002-07-03
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 3147
; TYPE: DNA
; ORGANISM: Homo Sapien
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (23)...(2589)
; OTHER INFORMATION: Nucleotide sequence encoding MTSP1
; PUBLICATION INFORMATION:
; AUTHORS: O'Brien, T.J. and Tanimoto, H.
; DATABASE ACCESSION NUMBER: GenBank #AR081724
; DATABASE ENTRY DATE: 2000-08-31
; PATENT DOCUMENT NUMBER: 5,972,616
; PATENT FILING DATE: 1998-02-20
; PUBLICATION DATE: 1999-10-26
US-10-190-030B-1
```

```
Query Match 60.6%; Score 1883.2; DB 13; Length 3147;
Best Local Similarity 81.2%; Pred. No. 0;
Matches 2223; Conservative 0; Mismatches 508; Indels 5; Gaps 3;

QY 45 GATCGGACCGCAAAACCATGGGTAGCAATCGGGGCGCAAGCGCGGAGGGGCTCTCAG 104
Db 5 GAGCGGCTCTGGGGTACCATGGGAGCGGATCGGGCGCCGCAAGGCGGAGGGGCGCCGAAG 64
QY 105 GACTTCGGCGCGGAGCTCAAGTACAACTCCCGGTAGAGAACATGATGGCTTTGAGGAG 164
Db 65 GACTTCGGCGCGGAGCTCAAGTACAACTCCCGGTAGAGAACATGATGGCTTTGAGGAG 124
QY 165 GGTGTGAGTTCCTGCTGCGAACCAATGCCAAGAAAGTGGAGAACGAGGCGCCAGCGCG 224
Db 125 GGCGTGGAGTTCCTGCTGCGAGTCAACAACTCAAGAGGTGGAAAGCATGGCGCGCGCGCG 184
```


Db 2345 ACCTGCGAGAACCTCTCTGCGCAGCAGATCACGCGCGCATGATGTGGTGGCTTCCTC 2404
Qy 2445 AGTGGGGTGTGGACTCTCTGCGAGGTGACATCTGTGTGCGCCCTTGTCAAGCGCGAGAA 2504
Db 2405 AGCGCGCGTGGACTCTCTGCGAGGTGATTCGGGGGACCCCTGTCCAGGTGGAGCG 2464
Qy 2505 GATGGCGGAATGTTCCAGGCTGGTGTGGTGAAGCTGGGTGAAGCTGGCTCAGAGAAC 2564
Db 2465 GATGGCGGAATCTTCCAGGCGGTGTGGTGAAGCTGGGTGAAGCTGGCTCAGAGAAC 2524
Qy 2565 AAGCCAGCGGTGTACACAGGCTCTCTGTAGTTGGGAGTGGATCAAGAGCACTGGG 2624
Db 2525 AAGCCAGCGGTGTACACAGGCTCTCTGTAGTTGGGAGTGGATCAAGAGCACTGGG 2584
Qy 2625 GTATAGCAGCATGACACAGCGCGACACAAACACACACAGGATGCGCCACATGCACA 2684
Db 2585 GTATAGGCGCGGG---GCCACCAATGTGTACCTGCGGGGCCACCCATCTCCACC 2641
Qy 2685 CCTGGATACAGGAGGAGAACACTGACGCAATTTATGTGTGGCTTCCCGCCCCCAACACA 2744
Db 2642 CCAGTGTGCAAG-CTGTGAGGCTGGAGACTGGACCGCTGACTGCACACAGCGCCC-CCAGA 2699
Qy 2745 ACCCAGACTGTGAATGCATCTTAGACTCAGAGT 2780
Db 2700 ACATACACTGTGAATCAATCTCCAGGCTCCAAAT 2735

RESULT 9
US-10-190-030B-3
; Sequence 3, Application US/10190030B
; Publication No. US20030134298A1
; GENERAL INFORMATION:
; APPLICANT: Madison, Edwin
; APPLICANT: Ong, Edgar
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING A
; TITLE OF INVENTION: TRANSMEMBRANE SERINE PROTEASE 20, THE ENCODED POLYPEPTIDES AND
; TITLE OF INVENTION: METHODS BASED THEREON
; FILE REFERENCE: 24745-1618
; CURRENT APPLICATION NUMBER: US/10/190.030B
; CURRENT FILING DATE: 2002-07-03
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 3147
; TYPE: DNA
; ORGANISM: Homo Sapien
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1865)...(2590)
; OTHER INFORMATION: Nucleic acid sequence of protease domain of MTSpl
US-10-190-030B-3

Query Match 60.6%; Score 1883.2; DB 13; Length 3147;
Best Local Similarity 81.2%; Pred. No. 0;
Matches 2223; Conservative 0; Mismatches 508; Indels 5; Gaps 3;

Qy 45 GATCGGACCGCCAAACCATGGTAGCAATCGGGCGCGCAAGCGCGGAGGGGCTCTCAG 104
Db 5 GAGCGGCTCGGGGTACCATGGGAGCGATCGGGCGCGCAAGCGCGGAGGGGCGCGGAG 64
Qy 105 GACTTCGGCGGGGACTCAAGTACAACTCCCGGCTAGAGACATGAATGGCTTTGAGAG 164
Db 65 GACTTCGGCGGGGACTCAAGTACAACTCCCGGCTAGAGACATGAATGGCTTTGAGAG 124
Qy 165 GGTGTGAGTTCCTGCTCGCAACAAATCCCAAGAAAGTGGAGAGCGGCGCCAGCGCG 224
Db 125 GCGGTGAGTTCCTGCTCGCAAGTCAAACTCAAGAGGTGGAAAGCATGGCCCCGGGCGC 184
Qy 225 TGGTGTGCTGGTGGAGTGTGTGAGTTCCTCTTGTCTCTCCCTCATGGCTGGCTTG 284
Db 185 TGGTGTGCTGGTGGAGTGTGTGAGTTCCTCTTGTCTCTCCCTCATGGCTGGCTTG 244
Qy 285 CTGGTGTGGCACTTCCATATTCGGAATGTGGGGTTCAAAAAGTCTTCAATGGCCATCTG 344

Db 245 CTGGTGTGGCAATTTGCAATACCGGACGTGCGGTGCCAGAGGTCTTCAATGGCTACATG 304
Qy 345 AGGATCAAAATGAGATCTTCTTGGATGCGTATGAGAACTCCACCTCCACAGATTTATC 404
Db 305 AGGATCAAAATGAGATTTTGTGGATGCGCTACGAGAACTCCAACTCCATGATTTGTA 364
Qy 405 AGCTGGCCAGCAGGTTGAAGGAGGCGCTCAAGCTGCTGTACAAATGAAGTCCCTGTCTG 464
Db 365 AGCTGGCCAGCAGGTTGAAGGAGGCGCTGAGCTGCTGTACAGCGAGTCCCATTTCTG 424
Qy 465 GGTCTCTTACCAAGAAAGTGGCTGTAACTGCTTCACTGAGGCGAGTGTATCGGCTAC 524
Db 425 GGGCCCTTACCAAGAGGTGCGCTGTGCGGCTTTCAGCGAGGCGCAGCTCATCGCTAC 484
Qy 525 TACTGTGTCAGAGTTTACAGATCCCGCCACACTGCGCAAGAGGTTGATCGGCGCATGGCT 584
Db 485 TACTGTGCTGAGTTTACAGATCCCGCCAGCACTGCTGTGAGGAGGCGCGCTCATGGCC 544
Qy 585 GTGGAGCGAGTTGTAACTTTGCCACCCCGGAGCAGCGGCACTGAAATCTTCTGTGCTAA 644
Db 545 GAGGAGCGCTGAGTGTGCTGCTGCGGCGGCTCCCTGAGTCTCTTGTGGTCACTC 604
Qy 645 TCTGTGTGCGCTTCCCATTTGACCCCGAGAAATGTGTCAGAGGACTCAGGCAACAGCTGC 704
Db 605 TCAGTGTGTGCTTCCCGCAGGACTCCAAACAGTACAGAGGACCCAGGCAACAGCTGC 664
Qy 705 AGTTTGTGCTGATGCCATGCTGTGAGCAGTGTGACGCTTCACTACCCCTGGCTTCCCT 764
Db 665 AGCTTTGGCTGTGACGCGCGGTGTGAGCTGATGCTTTCACGCGCGGCTTCCCT 724
Qy 765 AACAGTCCCTTACCGCGGCTATGCCGCTGTCAGTGGTCTCTGGGGGGGAGCGCGACTCT 824
Db 725 GACAGCCCTTACCGCGCTATGCCGCTGTCAGTGGGCTGCGGCGGAGCGCGACTCA 784
Qy 825 GTGCTGAGCTTCACTTCCGAGCTTTGATGTGCTCCCTGTGATGAGCATGAGGAGTGC 884
Db 785 GTGCTGAGCTTCACTTCCGAGCTTTGAGCTTTCGCTTTCGCGACGAGCGCGGAGCGAC 844
Qy 885 CTGGTCACTGCTGTATGATAGCTTGAAGCCCATGAGCAACCCACGCTGTGTGGCTGTGT 944
Db 845 CTGGTCACTGCTGTATGATAGCTTGAAGCCCATGAGCAACCCACGCTGTGTGGCTGTGT 904
Qy 945 GGCACCTTCTCACCTCTCTACACCTGACTTTCCTCTCCAGCAAGCTTCTTCCAGCTG 1004
Db 905 GGCACCTTACCTCTCTCTACACCTGACTTTCCTCTCCAGCAAGCTTCTTCCAGCTG 964
Qy 1005 ACCTGATTAACCAATGACTGACCGCGGAGCATCTTGGCTTTGAGGCGACTTCTTCCAGCTG 1064
Db 965 ACCTGATTAACCAATGACTGACCGCGGAGCATCTTGGCTTTGAGGCGACTTCTTCCAGCTG 1024
Qy 1065 CCAAGATGAGCAGCTGTGGCGGCTTTTGTGAGTGACACCCCAAGGAGCATTTAGCAGCCCC 1124
Db 1025 CCAAGATGAGCAGCTGTGGCGGCTTTGAGTGAGATCAATGAGGCGGAGCATTTCAACAGCCCC 1084
Qy 1125 TACTATCCAGGCGCACTACCGCGGCAACATCAATGACATGCAATGGAATATCAAGGTGCCAAC 1184
Db 1085 TACTATCCAGGCGCACTACCGCGGCAACATTTGATGCTGACATGGAATATTTAGGTGCCAAC 1144
Qy 1185 AACGGAGCTGAGGTGCGCTTCAAACTTCTTATCTGTGTGAGACCCCAAGTACCAAGTGT 1244
Db 1145 AACGAGCATGAGGTGAGCTTCAAACTTCTTATCTGTGTGAGACCCCAAGTACCAAGTGT 1204
Qy 1245 GGTCTCTGCAACGAGTATGAGGAGATCAAGGAGGAGTACTGCGGTGAGAGGTCC 1304
Db 1205 GGCACCTGCGGCGAGGACTGAGTGGAGATCAATGAGGAGGAGTACTGCGGAGAGAGGTCC 1264
Qy 1305 CAGTTGTGTGTGAGCAGCAACAGCAGCAGATTTACAGTCCACTTCCATTTGTATCATCTG 1364
Db 1265 CAGTTGTGTGTGAGCAGCAACAGCAGCAGATTTACAGTCCACTTCCACTCAGATCATCTG 1324
Qy 1365 TACAGCAGCAGCGGTTCTTAGTGTGAGTACTCTCTTACGACTCCCAAGCAGCCCGTCCCA 1424

Db 1325 TACACGACACCGGCTTCTTAGCTGAATACCTCTCTCTAGCTCCAGTCCAGCCATGCCCG 1384
QY 1425 GGGATGTTTCATGTGCAAGACTGGACGGTGCATCCGAAAGGAACTGGCTGCGACGGCTGG 1484
Db 1385 GGGCAGTTTCACTGCGGACACGGGGCGGTGTATCCGGAAGAGCTGGCTGTGATGGCTGG 1444
QY 1485 GCAGACTGCCGGATTTATAGTGTAGTGAAGCGTGTACTGCGGATGAATGCCACCCACCAATTC 1544
Db 1445 GCCGACTGACCGACACAGCGATGAGCTCAAATGTCAGTTGGACGCCCGCCACCAAGTTC 1504
QY 1545 ACGTCAAAAACCCAGTTCTGCAAGCCCTCTCTCTGGGCTCTGTGACAGTGTCAACGACTGT 1604
Db 1505 ACGTCAAGAAACAGTTCTGCAAGCCCTCTCTCTGGGCTCTGACAGTGTGAACGACTGC 1564
QY 1605 GGGACGGAAGTGAACGAGGAGGCTGACAGTGTCTGCTGGAGTTTCAAGTGTTCCAAT 1664
Db 1565 GGAGACAACAGCGACGAGCAGGGGTGACGTGTGCGGCCACAGCTTCAGGTGTTCAT 1624
QY 1665 GGGAGTGTCTCCCTCAGAGCCAGAGTGTATGGAAGGACAACTGTGGAGATGGGTCT 1724
Db 1625 GGGAGTGTCTCTCGAAAGCCAGCAGTGCATGGAAGGACGACTGTGGGACGGGTCC 1684
QY 1725 GACGAGCTTCATGTGACAGCGTGAATGTCTCTCTGACCAAAATATACCTACCGCTGC 1784
Db 1685 GACGAGCTCTCTGCCCAAGGTGAACGTCTGCTGTGACCAAAACACACCTACCGCTGC 1744
QY 1785 CAAATGGCTCTGTCTGACCAAGGCAACCTCTGAGTGTGATGGAAGACGAGCTGTAGC 1844
Db 1745 CTCAATGGCTCTGCTGTAGCAAGGCAACCTCTGAGTGTGACGGAAGGAGACTGTAGC 1804
QY 1845 GATGCTCCGATGAGNAACCTGTGACTGTGGCTGGATCTCTTACCAAAACAGGCTGCG 1904
Db 1805 GACGGCTCAGATGAGAAGGACTGCGACTGTGGCTGCGGTCTCAATTCACGAGACAGGCTGT 1864
QY 1905 GTGTTGTGGTGGCAAGTACGACGAGGCGAGTGGCTCTGAGTGTGATGGAAGACGAGCTGTAGC 1964
Db 1865 GTTGTGGGGCACAGTGCAGTGGAGGGGAGTGGCCCTGGAGTGAAGCTGATGTCT 1924
QY 1965 CTGGGCGAGGGCCACTGTGTGGGCTCGCTCATCTCTCTGACTGGCTGTCTGTCA 2024
Db 1925 CTGGGCGAGGGCCACTGTGGGTGCTTCCCTCATCTCTCCCAACTGTGGCTGTCTGTGC 1984
QY 2025 GCTCATGCTTTCAGGATGACAAAATTTCAAGTACTGAGCTACACGATGTGGACGGCC 2084
Db 1985 GCACACTGTATCATGATGACAGAGGATTCAGGTACTCAGACCCACGCGAGTGGACGGCC 2044
QY 2085 TTCTCTGGGTCTGTGACACAGAGCAAGCCAGTGTCTGGGTGACGAGCTGAAGCTC 2144
Db 2045 TTCTCTGGGTCTGACAGACAGAGCCAGCGAGGCCCTTGGGTGACGAGCGGAGCTC 2104
QY 2145 AAAAGTATCATCACCCACCTCTCTCAATGATTTTCACTTGCATGATGACATGCGCTTG 2204
Db 2105 AAGCGCATCATCTCCACCCCTCTTCAATGACTTCACTTGCATGATGACATGCGCTG 2164
QY 2205 CTGGAGCTGGAAGTCTGGTGGAGTACAGACCGTGTGCGGCCCATCTGCGCTGCTGAT 2264
Db 2165 CTGGAGCTGGAAGAACCGGACAGAGTACAGTCCATGTTGCGGCCCATCTGCTGCGGAC 2224
QY 2265 GCTACCCATGTCTTCTCTGTCGCAAGGCACTTGGGTGACAGGCTGGGGGCACACAAA 2324
Db 2225 GCTCCCATGTCTTCTCTGCGGCAAGGCCATCTGGGTGACGGGTGAGGACACACCCAG 2284
QY 2325 GAGGAGGTACCGGAGCGCTGATCTCTGCAAGAGGGGTGAGATCCGTGTCAATCAACAGACC 2384
Db 2285 TATGGAGGCACTGGGCGCTGATCTCTGCAAAAGGGGTGAGATCCGGGTCAATCAACAGACC 2344
QY 2385 ACCTGTGAGGACTCTGATCCGAGGAGATCAACCCAGATGATGTGTGGTTCCTC 2444
Db 2345 ACCTGGAGAACCTCTCTGCGGACAGAGATCAACCGCGCATGATGTGGGTGGGTTCCTC 2404
QY 2445 AGTGGGGGTGTGGAATCTCTGCGCAGGGGTGACTCTGTGTGGCCCTTGTCAAGCGCGGAGAA 2504
Db 2405 ACGCGGGGCTGGAATCTCTGCGCAGGGGTGATTCGGGGGAGCCCTGTCTCCAGCGTGGAGCG 2464

QY 2505 GATGGCGGAATTTTCCAGGCTGCTGTGTGTGAGCTGGGGTGAAGGCTGCGCTCAGAGGAAC 2564
Db 2465 GATGGCGGATTTTCCAGGCTGCTGTGTGTGAGCTGGGGAGACGGCTGCGCTCAGAGGAAC 2524
QY 2565 AAGCCAGGCGTGTACACAAGGCTCCCTGTGTGTGTGAGTTCGGGACTGGATCAAGAGACACTGGG 2624
Db 2525 AAGCCAGGCGTGTACACAAGGCTCCCTGTGTGTGTGAGTTCGGGACTGGATCAAGAGAACTGGG 2584
QY 2625 GTATAGCAGCTGACAGACAGCGACCGACCAACACCCACAGGATGCGCGACATGCACA 2684
Db 2585 GTATAGGCGCCCGG---GCCACCCAAATGTGTACACCTGCGGGGCGCACCATGTGTCACCC 2641
QY 2685 CTTGTGATACAGAGAGGAACACTGACGACATTTATGCTGTGGCTCTCCCGCCCAACACA 2744
Db 2642 CCAGTGTGACG-CCTGACGGCTGGAGACTGACCGCTGACTGACACCGAGGCC-CCAGA 2699
QY 2745 ACCGAGCTGTGAATGATGATCTTGGAGCTCAGAGT 2780
Db 2700 ACATACACTGTGAATCAATCTCCAGGGCTCCAAAT 2735

RESULT 10
US-10-267-219-1
; Sequence 1, Application US/10267219
; Publication No. US20030143219A1
; GENERAL INFORMATION:
; APPLICANT: Yeh, Jiunn-Chern
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING A TRANSMEMBRANE SERINE PROTEASE 2;
; TITLE OF INVENTION: ENCODED POLYPEPTIDES AND METHODS BASED THEREON
; FILE REFERENCE: 24745-1621
; CURRENT APPLICATION NUMBER: US/10/267,219
; PRIORITY FILING DATE: 2002-10-08
; PRIOR APPLICATION NUMBER: 60/328,530
; PRIORITY FILING DATE: 09-OCT-2001
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 3147
; TYPE: DNA
; ORGANISM: Homo Sapien
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (23)...(2589)
; OTHER INFORMATION: Nucleotide sequence encoding MTSP1
; PUBLICATION INFORMATION:
; AUTHORS: O'Brien, T.J. and Tanimoto, H.
; DATABASE ACCESSION NUMBER: GenBank #AR081724
; DATABASE ENTRY DATE: 2000-08-31
; PATENT DOCUMENT NUMBER: 5,972,616
; PATENT FILING DATE: 1998-02-20
; PUBLICATION DATE: 1999-10-26
US-10-267-219-1

Query Match 60.6%; Score 1883.2; DB 13; Length 3147;
Best Local Similarity 81.2%; Pred. No. 0;
Matches 2223; Conservative 0; Mismatches 508; Indels 5; Gaps 3;

QY 45 GATCGGACCGCCAAACCATGTTAGCAATCGGGCCCGCAAGCGCGAGGGGCTCTCAG 104
Db 5 GAGCGGCTCGGGGTACCATGTTGGAGCGATCGGGCCCGCAAGGGCGAGGGGCGCCGAG 64
QY 105 GACTTCGGCGCGGACTCAAGTACAATCTCCCGCTAGAGAACTGAATGGCTTTGAGGAG 164
Db 65 GACTTCGGCGCGGACTCAAGTACAATCTCCCGCTAGAGAACTGAATGGCTTTGAGGAG 124
QY 165 GGTGTGAGTTCCTGCTGCGAAACAATGCCAAGAAAGTGGAGAGCGGAGGCCCGAGCGCC 224
Db 125 GGGTGGAGTTTCTGCGCAGTCAACACGCTCAAGAGGTGGAAAGAGCATGGCCCGGGCGC 184
QY 225 TGGTGGTGTGCTGGTGGAGTGTCTGTGAGCTTCTCTTGTCTCTCCCTCATGGCTGGCTTG 284

QY 2445 AGTGGGGTGTGGACTCCTGCGCAGGGTGAATCTGTGGGCCCCCTGTGCAAGCCGAGAAA 2504
Db 2405 AGCGCGCGGTGGACTCCTGCGCAGGGTGAATCCGGGGACCCCTGTCCAGCGTGGAGCG 2464
QY 2505 GATGGCGGAATGTTCCAGGCTGGTGTGGTGAAGTGGGGTGAAGGCTGCGCTCAGAGGAAC 2564
Db 2465 GATGGCGGAATGTTCCAGGCGGTGTGGTGAAGTGGGGTGAAGGCTGCGCTCAGAGGAAC 2524
QY 2565 AAGCCAGCGGTGTACACAAGGCTCCCTGTAGTTCGGGACTGGATCAAGAGACACACTGGG 2624
Db 2525 AAGCCAGCGGTGTACACAAGGCTCCCTGTAGTTCGGGACTGGATCAAGAGAGAACTGGG 2584
QY 2625 GTATGACGACATGGACAGACAGCCACACACACACACACAGGATGCCGACATGCA 2684
Db 2585 GTATAGGGGGCGGG---GCCACCAAAATGTGTACACCTGCGGGGGCCACCCATCGTCCACC 2641
QY 2685 CCTGATACAGAGAGGAACACTGACGACATTTATGCTGTGGCTCCCTCCCGCCCAACACA 2744
Db 2642 CCAGTGTGACG-CCTGACGGCTGGAGACTGGACCGCTGACTGCACAGGGGCC-CCAGA 2699
QY 2745 ACCGAGCTGGAATGATCCTTTAGGACTCAGAT 2780
Db 2700 ACATACACTGTGAATCAATCTCCAGGGCTCCAAAT 2735

RESULT 11

US-10-267-219-3
; Sequence 3, Application US/10267219
; Publication No. US20030143219A1
; GENERAL INFORMATION:
; APPLICANT: Madison, Edwin
; APPLICANT: Yeh, Jium-Chern
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING A TRANSMEMBRANE SERINE PROTEASE 2
; FILE OF INVENTION: ENCODED POLYPEPTIDES AND METHODS BASED THEREON
; FILE REFERENCE: 24745-1621
; CURRENT APPLICATION NUMBER: US/10/267,219
; PRIOR FILING DATE: 2002-10-08
; PRIOR APPLICATION NUMBER: 60/328,530
; PRIOR FILING DATE: 09-OCT-2001
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 3147
; TYPE: DNA
; ORGANISM: Homo Sapien
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1865)...(2590)
; OTHER INFORMATION: Nucleic acid sequence of protease domain of MTSPI
US-10-267-219-3

Query Match 60.6%; Score 1883.2; DB 13; Length 3147;
Best Local Similarity 81.2%; Pred. No. 0;
Matches 2223; Conservative 0; Mismatches 508; Indels 5; Gaps 3;
QY 45 GATCGACCGCCAAACCATGGTAGCAATCGGGCCCGCAGGCGCGAGGGGCTCTCAG 104
Db 5 GAGCGGCTCGGGGTACCATGGGAGCGATCGGGCCCGCAAGGGCGAGGGGCGCCGAAG 64
QY 105 GACTTCGGCGCGGGACTCAAGTACAACCTCCCGGCTAGAGAACATGAATGGCTTTGAGGAG 164
Db 65 GACTTCGGCGGGGACTCAAGTACAACCTCCCGGCGAGGAAAGTGAATGGCTTTGAGGAA 124
QY 165 GGTGTGGAGTCTTCCTGCGTGGCAACAAATCCAGAAAGTGGAGAGCGAGGCCCGCAGGCGC 224
Db 125 GCGGTGGAGTCTTCCTGCCAGTCAACAACTGCAAGAGGTGGAAAGCATGGCCCGGGGCGC 184
QY 225 TGGGTGGTGTGGTGGAGTGTCTGAGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 284
Db 185 TGGGTGGTGTGGTGGAGCGGTGTCTGATCGGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 244
QY 285 CTGGTGTGGCACTTCCATATATCGGAATCTGGGGTTCAAAAGTCTTCAATGGCCATCTG 344

Db 245 CTGGTGTGGCACTTTCAGTACCGGGACGTGGGTGTCCAGAAAGGTCTTCAATGGCTACATG 304
QY 345 AGGATCACAAAATGAGATCTTTCTGGATCGGTATGAGAACTCCACCTCCACAGATTTATC 404
Db 305 AGGATCACAAAATGAGAAATTTTGTGGATGCTTACGAGAACTCCAACTCCACTGAGTTTGA 364
QY 405 AGCCTGGCCAGCCAGGTGAAGGAGGCGCTGAAGCTGTGTACAAATGAATGCTCCTCTCTCTG 464
Db 365 AGCCTGGCCAGCCAGGTGAAGGAGGCGCTGAAGCTGTGTACAGCGGAGTCCCACTTCCTG 424
QY 465 GGTCCCTACCAACAAGAAGTCTGGCTGTAACTGGCTTCAGTGGAGGAGTGTCTATCCGCTAC 524
Db 425 GGGCCCTTACCAACAAGGAGTCTGGCTGTGACGGCTTTCAGCGGGGAGGCTCATCGCTAC 484
QY 525 TACTGGTTCAGAGTTTCAGCATCCCCCACACCTGGCAGAGAGTTGTATCGGCCATGGCT 584
Db 485 TACTGGTTCAGAGTTTCAGCATCCCCCACACCTGGCAGAGAGTGTGTGGAGGAGCGGCTATGGCC 544
QY 585 GTGAGGAGGAGTTGTAACATTTCCACCCCGCAGCAGCGGCACTGAAATCTCTTGTGTAAACA 644
Db 545 GAGGAGGCGGTAGTCACTGTCTGCCCGCGGGCGCTCCCTGAAAGTCTCTTGTGTGTCACC 604
QY 645 TCTGTGGTGGCTTTCCCAATTCAGCCAGAAATGTGTCAGAGAGATCTCAGAGCAACAGCTGC 704
Db 605 TCAGTGGTGGCTTTTCCCAACCGGACTCCAAAACAGTACAGAGGAGCCCAAGCAACAGCTGC 664
QY 705 AGTTTTCGCCCTGCACTGCTGTCAGCAGTACACGCTTCACTTACCCCTGGCTTCCCC 764
Db 665 AGCTTTCGCCCTGCACTGCTGTCAGCAGTGTGGAGTGTATGCGCTTCAACACGCCCGGCTTCCCT 724
QY 765 AACAGTCTCTACCCGGCGCATGCGCGTGTCCAGTGGGTCTCTGGCGGGGAGACCGCACTCT 824
Db 725 GACAGCCCTTACCCCGCTCATGCGCGTGTCCAGTGGGCGCTTCCGCGGGGAGACCGCACTCA 784
QY 825 GTGCTGAGCCTCACTTCCGAGCTTGTATGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 884
Db 785 GTGCTGAGCCTCACTTCCGAGCTTGTACCTTGGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 844
QY 885 CTGGTCACTGCTATGATGATGCTGAGCCCACTGAAACCCCAAGCTGTGGTGGGCTGTGT 944
Db 845 CTGGTCACTGCTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 904
QY 945 GGCACCTTCTCACCTCTCTCAACCTGACTTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1004
Db 905 GGCACCTTCTCACCTCTCTCAACCTGACTTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 964
QY 1005 ACGCTGATAACCAATACTGACCGGGGAGATCTGCTGGCTTGGAGGCACTTTCTCTCTCTCT 1064
Db 965 AACTGATAACCAACACTGAGCGGGGAGATCTGCTGGCTTGGAGGCACTTTCTCTCTCTCT 1024
QY 1065 CCCAAGATGAGCAGCTGTGGCGGCTTTTGTAGTGACACCCCAAGGGAGCAATTTAGCAGCCCC 1124
Db 1025 CCTAGGATGAGCAGCTGTGGAGGCGGCTTACGTAAAGCCCAAGGGGAGCAATTCACAGCCCC 1084
QY 1125 TACTATCCAGGCACTACCGCCCAACATCACTGACATGGAATATCAAGGTGCCCAAC 1184
Db 1085 TACTATCCAGGCACTACCGCCCAACATCACTGACATGGAATATCAAGGTGCCCAAC 1144
QY 1185 AACCGGAACGTGAGGTCGCTTCAAACTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1244
Db 1145 AACCGATGTGAAGGTGAGCTTCAAAATTTCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1204
QY 1245 GGTCTCTGCAACAGGACTATGTGGAGATCAACGGGGGAGAAAGTACTCTGGGTGAGAGTCC 1304
Db 1205 GGCACCTGCCCCAAGGACTACGTGGAGATCAATGGGGAGAAATACTCGGGAGAGAGTCC 1264
QY 1305 CAGTTTGTGTGAGCAGAAACAGCAGCAAGTATACGTCCACTTCTCTCTCTCTCTCTCTCTCT 1364
Db 1265 CAGTTTGTGTGAGCAGAAACAGCAGCAAGTATACGTCCACTTCTCTCTCTCTCTCTCTCTCT 1324
QY 1365 TACAGGACACCGGTTCTTAGCTGAGTACTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1424
Db 1325 TACAGGACACCGGTTCTTAGCTGAGTACTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1384

QY	1425	GGGATGTTCTATGTGTCAGAGCTGGACGGTGCATCCGAAGGAACATGCGCTCGACGCGCTGG	1488
Db	1385	GGGCAGTTACGTGCGCCGACGGGCGGTGTATCCGGAAGGAGCTGCGCTGTGATGGGTGG	1444
QY	1485	GCAGACTGCCCGGANTTATGTGATGAGCGTTACTGCCGATGCAAATGCCACCCACCACTTC	1544
Db	1445	GCCGACTGCAACCGACACAGCGATGAGCTCAACTGSCAGTTTGCACGCGCGGCCACCACTTC	1504
QY	1545	ACGTGCAAAAACCAAGTTCTGTCGAAGCCCTCTTCTGGGTCTGTGACAGTGTCAACGACTGT	1604
Db	1505	ACGTGCAAGAACAGTTCTGCAAGCCCTCTTCTGGGTCTGTGACAGTGTGACGACTGC	1564
QY	1605	GGGACCGGAAGTGACGAGGAGGGCTGCAGCTGTCTCTCTGGAGTTTCAAGTGTTCAAAT	1664
Db	1565	GGAGACAACAGCGACGACGGGTGCAGTTGTCCGGCCCGACAGCTTCAGGTGTTCAAAT	1624
QY	1665	GGGAAGTGTCTCCCTCAGAGCCAGCAAGTGTAAATGGGAAGGACAACTCTGGAGATGGGTCT	1724
Db	1625	GGGAAGTGCCTCTCGAANAAGCCAGCTGCNAATGGGAAGACGACTGTGGGACGGGTCC	1684
QY	1725	GACGAGCCTCATGTGACAGCGTGAATGTCTGTCTCTTTCGACCAATATACCTACCGCTGC	1784
Db	1685	GACGAGCCTCCTGCCCAAAGTGAAAGCTGCTCACTTGTACCAACACACCTACCGCTGC	1744
QY	1785	CAAAATGSCCTCTGTGACGACGAGGCAACCTGAGTGTGATGGGAAGACGACTGTAGC	1844
Db	1745	CTCAATGGGCTCTGTCTTGACAAAGGCAACCCGTAGTGTGACGGGAAGGAGACTGTAGC	1804
QY	1845	GATGGCTCCGATGAGAAAAAATGTGACTGTGGGCTGCGATCCTTTACCAACACAGGCTCGC	1904
Db	1805	GACGGCTCAGATGAGAAGGACTGCGACTGTGGGCTGCGGTCACTTACGACGACAGGCTCGT	1864
QY	1905	GTGGTGGTGGCAGANTGCGHACGAGGCGAGTGGCCCTTGGCAGTGAAGCTCCAGGCC	1964
Db	1865	GTGTGTGGGGCACGATGCGGATGAGGGCGAGTGGCCCTTGGCAGGTAAAGCTTCGATGCT	1924
QY	1965	CTGGGCGAGGGCCACTTGTGTGGGGCCCTGCTCATCTCTCTGACTGGCTGGTCTCTGCA	2024
Db	1925	CTGGGCGAGGGCCACATCTCGGTGCTTCCTCATCTCTCCAACTGGCTGGTCTCTGCC	1984
QY	2025	GCTCATTTCTTTCAGATGACAAAAATTTCAAGTACTCAGACTACGATGTGGAAGGCC	2084
Db	1985	GCACACTGCTACATCGATGATGACAGAGGATTCAGGTACTTCAGACCCACGCGAGTGGACGCC	2044
QY	2085	TTCTCTGGGTCTGTGACACAGAGCAAGCGCAGTGCCTCTCGGGTGCAGGAGCTGAAGCTC	2144
Db	2045	TTCTCTGGGCTTGACGACACAGAGCCAGCGCGAGCGCCCTGGGGTGCAGGAGCGCAGGCTC	2104
QY	2145	AAACGTATCATACCCACCTTCTCTTCAATGATTTCACTCTTCAAGTATGATATGCGCTTG	2204
Db	2105	AAGCGCATCATCTCCACCCCTTCTTCAATGACTTCACTCTTCGACTATGATCATCGGCTG	2164
QY	2205	CTGGAGCTGAGAAAGTCGGTGGAGTACAGCACCGTCTGTGCGCCCATCTCTGCTGCTGAT	2264
Db	2165	CTGGAGCTGAGAAAACGGGCAGAGTACAGCTCCATGTTGCGGCCCATCTGCTTCCGCGAC	2224
QY	2265	GCTTACCATGTCTTTCCTGTGTGGAAGGCCATCTGGGTTCAGGGTGGGGGCAACAAAA	2324
Db	2225	GCCTCCCATGTCTTTCCTGTGCGGCAAGGCCCATCTGGGTTCAGGGTGGGGGACACACCCAG	2284
QY	2325	GAGGGAGGTACGGAGGCTGATCTTGCAAGAGGTGAGATCCGTGTCTATCAACCGAGCC	2384
Db	2285	TATGGAGGCACTGGCGGCTGTATCTTGCAAAGGGTGAGATCCGGTTCATCAACAGACC	2344
QY	2385	ACCTGTGAGGACCTCATTCGCGACGAGATCACCCCAAGATGATGTGTGTGGGTGTCTTCTC	2444
Db	2345	ACCTGCGAGAACTCTCTCGCGACGAGATCACGCCGCGCATGATGTGCTGTGGCTTCTTCTC	2404
QY	2445	AGTGGGGTGTGACTCTCTCGAGGCTGACTCTGTGTGGCCCTTGTCAAGCGCGGAGAAA	2504
Db	2405	AGCGGCGGCTGACTCTCTGCAAGGTTGATTTCCGGGGGACCCCTGTCCAGCTGTGAGGCG	2464

QY	2505	GATGGCGAATGTTCCAGGCTGGTGTGGTGGGCTGAAGGCTCGCTCAGAGAAC	2564
Db	2465	GATGGGCGGATCTTCCAGGGCGGTGTGTGAGCTGGGAGACGGCTCGCTCAGAGAAC	2524
QY	2565	AAGCCAGCGGTGTACACAAGGCTCCCTGTAGTTTCGGGACTGGATCAAGAGCACACTGGG	2624
Db	2525	AAGCCAGCGGTGTACACAAGGCTCCCTCTGTTCGGGACTGGATCAAGAGCACACTGGG	2584
QY	2625	GTATAGCAGCATGGACAGACAGCGGACCAAAACACCCACAGGAGTCCCGCATGCACA	2684
Db	2585	GTATAGGGGGCGGG--GCCACCCAAATGTGTACACCTGGCGGGGCCACCCATCGTCCACC	2644
QY	2685	CCTGGATACAGGAGGAGAACACTGACGACATTTATGCTGTGGCTCTCCCGCCCAACACA	2744
Db	2644	CCATGTGTGACG--CCTGAGGCTGGAGACTGGACCGCTGACTGCACCGGCCC--CCAGA	2699
QY	2745	ACCAGACTGTGAACCTCATCTTTAGGACTCAGAGT	2780
Db	2700	ACATACACTGTGAACCTCAATCTCCAGGGCTCCAAAT	2735

RESULT 12

US-10-302-840A-1

; Sequence 1, Application US/10302840A

; Publication No. US20030134794A1

; GENERAL INFORMATION:

; APPLICANT: Madison, Edwin L.

; APPLICANT: Ong, Edgar O.

; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING SERINE PROTEASE CVSP17, A

; TITLE OF INVENTION: POLYPEPTIDES AND METHODS BASED THEREON

; FILE REFERENCE: 24745-1622

; CURRENT APPLICATION NUMBER: US/10/302,840A

; CURRENT FILING DATE: 2003-01-24

; PRIOR APPLICATION NUMBER: 60/332,015

; PRIOR FILING DATE: 2001-11-20

; NUMBER OF SEQ ID NOS: 18

; SOFTWARE: PastSeq for Windows Version 4.0

; SEQ ID NO 1

; LENGTH: 3147

; TYPE: DNA

; ORGANISM: Homo Sapien

; FEATURE:

; NAME/KEY: CDS

; LOCATION: (23)...(2589)

; OTHER INFORMATION: Nucleotide sequence encoding MTSPI

; PUBLICATION INFORMATION:

; DATABASE ACCESSION NUMBER: GenBank #AR081724

; DATABASE ENTRY DATE: 2000-08-31

US-10-302-840A-1

Query Match	60.6%;	Score 1883.2;	DB 13;	Length 3147;
Best Local Similarity	81.2%;	Pred. No. 0;		
Matches 222;	Conservative	0;	Mismatches 508;	Indels 5; Gaps 3;
Qy	45	GATCGNACCGCCAAAACCATGGGTAGCAATCGGGGCGGAAGCCGAGGGGGCTCTCAG	104	
Db	5	GAGCGGCTCTGGGGTACCATGGGAGCGATCGGGCCGCGAAGCGGAGGGGGCCCGAAG	64	
Qy	105	GACTTCGGCGCGGGACTCAAGTCAACTCCCGCTAGAGAACAATGAATGCTTTGAGAG	164	
Db	65	GACTTCGGCGCGGGACTCAAGTCAACTCCCGCGACGAGAAAGTGAATGCTTTGGAGGAA	124	
Qy	165	GGTGTGGAGTTCCTGCTCGCAACAATGCCAAGATGCAGAAAGCGAGGCCCCACGCGC	224	
Db	125	GGCGTGGAGTTCCTGCCAGTCAACAGTCAAGAGGTGAAAAAGCATGCGCCCGGGCGC	184	
Qy	225	TGGGTGGTGCTGTGGCAGTGCTGTTTCAGTCTCTTGCTCTCCCTCATGSGGTGGCTTG	284	
Db	185	TGGGTGGTGCTGTGCAGCGGTGCTGATCGGGCTCTCTTGCTCTGTGGGATCGGCTTC	244	
Qy	285	CTGGTGGGCATTCGATATCGGAATGTGCGGGTCAAAAAGTCTTCAATGCCCATCTG	344	
Db	245	CTGGTGGCATTTGACGTACCGGGACGTGCGGTGCCAAGGCTCTCAATGCTACATG	304	

345 AGGATCAAAATGAGATCTTTCTGATCGGTATGAGAACTCCACCTCCACAGAGTTTATC 404
Db |||||
305 AGGATCAAAATGAGAAATTTTGTGATCGCTACGAGAACTCCAACTCCACTGAGTTTGT 364
Qy |||||
405 AGCTGGCCAGCCAGGTGAGAGAGCGCTGAGCTGCTGTGTACAAATGAAGTCCCTGCTTG 464
Db |||||
365 AGCCTGGCCAGCAAGGTGAAGACCGCTGAAAGCTGCTGTGTACAGCGGAGTCCCAATTCCTG 424
Qy |||||
465 GGTCCCTACCAAGAAAGTGGCTGTAACTGCTTTCACTGAGGGCAGTGTCTATGCCCTAC 524
Db |||||
425 GGCCTTACCAAGAGTGGCTGTGACGGCTTTACGGAGGGCAGCTCATGCCCTAC 484
Qy |||||
525 TACTGGTCAAGTTTCAAGATCCCCCAACCTGGCAGAGAGTTGTATCGGCCATGGCT 584
Db |||||
485 TACTGGTCTGAGTTTCAAGATCCCGCAGCACTGTGTGAGGAGGGCCGAGCGGTCTATGCC 544
Qy |||||
585 GTGAGCGAGTTGTAACATTGCCACCCGAGCAGCGGCACTGAATCTCTTGTGTAAACA 644
Db |||||
545 GAGGAGCGGTAGTCACTGCTGCCCGCGGGCGCTCCCTGAAGTCTCTTTGTGTACCC 604
Qy |||||
645 TGTGTGGTGGCTTCCCAATTGACCCAGAAATGCTGCAGAGGACTCAGAGCAAAAGCTGC 704
Db |||||
605 TCAGTGGTGGCTTTCCCAAGCACTCCAAACAGTACAGAGGACCCAGAGCAACAGCTGC 664
Qy |||||
705 AGTTTGGCTGCATGCCCCATGTTGAGAGTGACACAGCTTCACTAACCCCTGGCTTCCCC 764
Db |||||
665 AGCTTTGGCTGCACGCCCGCGGTGTGAGCTGATGCGCTTCAACCACGCCCGGCTTCCCT 724
Qy |||||
765 AACGTCTCCATCCGGCGCATGCGCTGCCAGTGGGTCTCTGGGGGAGCGCGACTCT 824
Db |||||
725 GACAGCCCTTACCCCGCTCATGCGCTGCCAGTGGGGCCCTGCGGGGGAGCGCGACTCA 784
Qy |||||
825 GTGCTGAGCTTCACTTCCGAGCTTTGATGCTGCTTCCCTGTGATGAGCATGGCAGTGC 884
Db |||||
785 GTGCTGAGCTTCACTTCCGAGCTTTGACCTTGGTCTCTGCGAGCGCGCGAGCGAC 844
Qy |||||
885 CTGGTCAACGTTGATGATGAGCTTGAAGCTTGAAGCTTGAAGCTTGAAGCTTGAAGCT 944
Db |||||
845 CTGGTCAACGTTGATGATGAGCTTGAAGCTTGAAGCTTGAAGCTTGAAGCTTGAAGCT 904
Qy |||||
945 GGCACCTTCTCAACCTGACTTTCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1004
Db |||||
905 GGCACCTTCT 964
Qy |||||
1005 ACGTGTATACCAATPACTGACCGCGCATCTGCTGCTTGAAGGCACTTCTTCCAGCTG 1064
Db |||||
965 ACCTGATACCAACACTGAGCGCGCATCCCGCTTGAAGGCACTTCTTCCAGCTG 1024
Qy |||||
1065 CCCAGATGAGAGTGTGGCGCTTTTGTGATGACACCCAGGAGCAATTTAGCAGCCCC 1124
Db |||||
1025 CCTAGGATGACAGCTGTGGAGCGCTTACGTTAAAGCCAGGAGCAATTTCAAGCCCC 1084
Qy |||||
1125 TACTATCCAGCCCACTACCCGCCAACAATCACTGCATGATATCAAGTGGCCCAAC 1184
Db |||||
1085 TACTATCCAGCCCACTACCCGCCAACAATCACTGCATGATATCAAGTGGCCCAAC 1144
Qy |||||
1185 AACCGAACTGAAGGTGCTTCAAACTTCTTATCTGTGGAGCCCAAGTACCACTAGT 1244
Db |||||
1145 AACAGCATGTGAAGTGTGAGCTTCAAACTTCTTATCTGTGGAGCCCAAGTACCACTAGT 1204
Qy |||||
1245 GGCTCTGCAACAGGACTATGTGAGATCAACGGGGAGAGTACTGGGGTGAAGTTC 1304
Db |||||
1205 GGCACCTCCCCAAGGACTAGCTGGAGATCAATGGGGAGAAATACTGGGGAGAGAGTTC 1264
Qy |||||
1305 CAGTTTGTGTGAGCAGCAACAGCAGCAAGATTAAGTTCACCTTCCATCTTGTATCACTCG 1364
Db |||||
1265 CAGTTTGTGTGAGCAGCAACAGCAGCAAGATTAAGTTCACCTTCCATCTTGTATCACTCG 1324
Qy |||||
1365 TACAGGACACCGGGTTCCTTAGCTGAGTACCTCTCTTACGACTTCAACAGCAGCCGTGCCA 1424
Db |||||
1325 TACACCGACACCGGGTTCCTTAGCTGAGTACCTCTCTTACGACTTCAACAGTACCAATGCCG 1384

1425 GGGATGTTTCATGTGCAGACTGGACGGTGCATCCGAAAGGAACTCGCTGCGACCGCTGG 1484
Db |||||
1385 GGGCAGTTTTCATGTGCAGACTGGACGGTGCATCCGAAAGGAACTCGCTGCTGTGATGGCTG 1444
Qy |||||
1485 GCAGACTGCCCGGATATAGTATGATGAGGCTTACTGCCGATGCAATGCCACCCAGTTTC 1544
Db |||||
1445 GCCAGCTGCACCGACCAAGAGGATGAGCTCAACTGCACTGCGAGCGCGGCGCCAGTTTC 1504
Qy |||||
1545 ACCTGCAAAACCAAGTTCTGCAAGCCCTCTTCTGGGTCTGTGACAGTGTCAACAGCTGT 1604
Db |||||
1505 ACCTGCAAAACCAAGTTCTGCAAGCCCTCTTCTGGGTCTGTGACAGTGTCAACAGCTGT 1564
Qy |||||
1605 GGGAGCGGAAGTGCAGAGGGGCTGCAGCTGTCTGCTGGGAGTTTCAAGTGTTCCTCAAT 1664
Db |||||
1565 GGGAGCAACAGCAGCAGCAGGAGTGTGCTGCGGCCAGACCTTCAGGTGTTCCTCAAT 1624
Qy |||||
1665 GGGAGTGTCTCTCCCTCAGAGCCAGAGTGTAAATGGGAAGGAACTGTGAGAGATGGGTCT 1724
Db |||||
1625 GGGAGTGTCTCTCGAAAGCCAGCAGTGTCAATGGGAAGGAGCACTGTGGGAGCGGTTC 1684
Qy |||||
1725 GACGAGGCTTTCATGTGCAGCGTGAATGTCTCTCTTGCACCAAAATATACCTACCGCTGC 1784
Db |||||
1685 GACGAGGCTTCTGCTGCCCAAGTGTGAAGTGTCACTTGTACAAACACACCTTACCGCTGC 1744
Qy |||||
1785 CAAATGGCTCTCTGCTGAGCAAGGCAACCTGAGTGTGATGGGAAGCAGGACTGTAGC 1844
Db |||||
1745 CTCATGGCTCTGCTTGGAGCAAGGCAACCTGAGTGTGACGCGGAGGAGGACTGTAGC 1804
Qy |||||
1845 GATGGCTCCGATGAGAAACCTGTGACGTTGGGTGCGATCTTTTACAAACAGGCTCGC 1904
Db |||||
1805 GACGCTCAGATGAGAAAGGACTGCGACTGTGGGTGCGGTCAITTCACGAGACAGGCTCGT 1864
Qy |||||
1905 GTGCTTGTGTGCACAAATGCGGACGAGGCGAGTGGCTTGGCAGGTCGAGCTCCACGCC 1964
Db |||||
1865 GTTGTGGGGCAGAGATGCGGATGAGGCGAGTGGCCCTGCGAGGTAAAGCTGCTGCTG 1924
Qy |||||
1965 CTGGGCGAGGCGCTCTGCTGAGGCGCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTGCA 2024
Db |||||
1925 CTGGGCGAGGCGCTCTGCTGAGTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTG 1984
Qy |||||
2025 GCTCATTTGCTTTCAGGATGACAAAAATTTCAAGTACTCAGACTACACGATGTGACGCGC 2084
Db |||||
1985 GCACACTGCTACATCGATGACAGAGGATTCAGGTACTCAGACCCCAACGAGTGTGACGCGC 2044
Qy |||||
2085 TTCTGCTGCTGCTGGACAGCAAGCGAGTGCCTCTGCGGTGCGAGGCTCAAGCTC 2144
Db |||||
2045 TTCTGCTGCTGCTGACAGCAGCAGCAGCGCGCGCCCTGCGGTGCGAGGCGAGGCTC 2104
Qy |||||
2145 AAAGTATCATCACCCACCTTCTCTCAATGATTTCACTTTCGACTATGACATCGCCTTG 2204
Db |||||
2105 AAGCGCATCATCTCCACCCCTTCTTCAATGACTTCACTTTCGACTATGACATCGCGCTG 2164
Qy |||||
2205 CTGAGCTGAGAGTGTGAGTACAGCACGCTGCTGCGCCCACTCTGCTGCTGCTGAT 2264
Db |||||
2165 CTGAGCTGAGAGTGTGAGTACAGCTTCAATGCTGCGGCGCCCTTCTGCTGCGGAC 2224
Qy |||||
2265 GCTACCATCTCTCTCTGCTGCAAGGCGCTTGGGTTCAGAGGCTGGGGGACACACAAA 2324
Db |||||
2225 GCCTCCCATCTCTCTCTGCGGCAAGGCGCTTGGGTTCAGAGGCTGGGGACACACCCAG 2284
Qy |||||
2325 GAGGAGGTACCGAGCGCTGATCTCTGAGAGGCTGAGTTCCTGATCAACACAGGAC 2384
Db |||||
2285 TATGAGGCGCTGCGCGCTGATCTCTGCAAAAGGCTGAGATCCGCGTCACTCAACAGGAC 2344
Qy |||||
2385 ACCTGTGAGGACCTCATGCGCAGCAGATCAACCCCAAGATGATGTGTGGGTTCCTC 2444
Db |||||
2345 ACCTGTGAGGACCTCTCTGCGCAGCAGATCAACCGCGCGATGATGTGTGGGTTCCTC 2404
Qy |||||
2445 AGTGGGCTGTGACTCTCTGCGAGGCTGACTCTGCTGGTGGCGCTTGTCAAGCGCGGAGAA 2504
Db |||||
2405 AGCGGCGGCTGAGCTCTCTGCGAGGCTGATTCGGGGGACCCCTGTCAAGGCTGAGGCG 2464
Qy |||||
2505 GATGGGCGAATGTTTCCAGGCTGTGTGTGAGCTGGGGTGAAGGCTGCGCTCAGAGGAC 2564

```
Db 2465 GATGGCGGATCTTCCAGGCGGCTGTGGTGAAGTGGGAGACGGCTGCGCTCAGAGAAC 2524
Qy 2565 AAGCCAGCGGTGTACAAAGGCTCCCTGTAGTTCGGGACTGGATCAAGAGACACTGGG 2624
Db 2525 AAGCCAGCGGTGTACAAAGGCTCCCTGTAGTTCGGGACTGGATCAAGAGACACTGGG 2584
Qy 2625 GTATAGCAGCATGACAGACAGCGGACCAAAACACCCACAGGAGATCCCGACATGCACA 2684
Db 2585 GTATAGGCGCGGG---GCCACCCAAATGTGTACACCTGCGGGCCACCCCATCGTCCACC 2641
Qy 2685 CCTGGATACAGGAGGAGGACACTGACGACATTTATGCTGTGGCTCCCGCCCCCAACACA 2744
Db 2642 CCAATGTGCAGC-CCTGAGGCTGGAGACTGGACCGGTGACTGCACCAAGCGCCC-CCAGA 2699
Qy 2745 ACCAGACTGTGAATGCTCATCTTAGGACTCAGAGT 2780
Db 2700 ACATACATGTGAATCAATCTCCAGGGCTCAAT 2735

RESULT 13
US-10-302-840A-3
; Sequence 3, Application US/10302840A
; Publication No. US20030134794A1
; GENERAL INFORMATION:
; APPLICANT: Madison, Edwin L.
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING SERINE PROTEASE CVSP17, THE ENCOD
; TITLE OF INVENTION: POLYPEPTIDES AND METHODS BASED THEREON
; FILE REFERENCE: 24745-1622
; CURRENT APPLICATION NUMBER: US/10/302,840A
; CURRENT FILING DATE: 2003-01-24
; PRIOR APPLICATION NUMBER: 60/332,015
; PRIOR FILING DATE: 2001-11-20
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 3147
; TYPE: DNA
; ORGANISM: Homo Sapien
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1865)...(2590)
; OTHER INFORMATION: Nucleic acid sequence of protease domain of MTSPL
US-10-302-840A-3

Query Match 60.6%; Score 1883.2; DB 13; Length 3147;
Best Local Similarity 81.2%; Pred. No. 0;
Matches 2223; Conservative 0; Mismatches 508; Indels 5; Gaps 3;

Qy 45 GATCGGACCGCCAAACCATGGGTAGCAATCGGGGCGCGAAGCGCGGAGGGGCTCTCAG 104
Db 5 GAGCGGCTCGGGGTACCATGGGAGGCGATCGGGCCCGCAAGGGCGGAGGGGCGCGAAG 64
Qy 105 GACTTCGGCGGGGACTCAAGTACAACTCCCGGTAGAGAACATGAATGGCTTTGAGGAG 164
Db 65 GACTTCGGCGGGGACTCAAGTACAACTCCCGGTAGAGAACATGAATGGCTTTGAGGAG 124
Qy 165 GGTGTGAGTTCCTGCTCGGAAACATGCCAAGAAAGTGGAGAAAGCGAGGCCCGCGAGCGC 224
Db 125 GGCCTGGAGTTCTCGCCAGTCAACAAAGTCAAGAGGTGGAAAGCATGCGCCGGGCGC 184
Qy 225 TGGGTGTGCTGTGGGAGTGTGCTGTAGCTTCCTCTTGTCTCTCCCTCATGCGTGGCTTG 284
Db 185 TGGGTGTGCTGTGGGAGCGGTGCTGATGGGCTCTCTTGTGCTTGTGGGATCGGCTTC 244
Qy 285 CTGTGTGGCACTTCCATATTCGAAATGTGCGGTTCAAAAAGTCTCAATGGCCATCTG 344
Db 245 CTGTGTGGCAATTTGAGTACCGGAAGTGTGCTGCTCAGAGGTCTTCAATGGCTCATG 304
Qy 345 AGGATCAAAATGAGATCTTTCTGGATGCGGTATGAGAACTCCACTCCACAGAGTTTATC 404
Db 305 AGGATCAAAATGAGATTTTGTGGATGCGCTTACGAGAACTCCAACTCCACTGAGTTGTA 364
```

```
Qy 405 AGCTGTCCAGCCAGGTGAAGGAGGCGCTCAAGTCTGTGTACAAATGAAGTCCCTGTCTGTG 464
Db 365 AGCTGTCCAGCCAGGTGAAGGAGGCGCTCAAGTCTGTGTACAAATGAAGTCCCTGTCTGTG 424
Qy 465 GGTCCCTTACCAAGAAAGTGGGTGTAACTGCTTCACTGAGGCGAGTGTATCGGCTAC 524
Db 425 GGCCTTACCAAGAAAGTGGGTGTAACTGCTTCACTGAGGCGAGTGTATCGGCTAC 484
Qy 525 TACTGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGT 584
Db 485 TACTGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGT 544
Qy 585 GTGGAGGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGT 644
Db 545 GAGGAGGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGT 604
Qy 645 TCTGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGT 704
Db 605 TCAAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGT 664
Qy 705 AGTTTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGT 764
Db 665 AGTTTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGT 724
Qy 765 AACAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGT 824
Db 725 GACAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGT 784
Qy 825 GTGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGT 884
Db 785 GTGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGT 844
Qy 885 CTGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGT 944
Db 845 CTGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGT 904
Qy 945 GGCAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGT 1004
Db 905 GGCAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGT 964
Qy 1005 ACCTGTGATTAACCAATGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 1064
Db 965 ACCTGTGATTAACCAATGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 1024
Qy 1065 CCAAGATGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGT 1124
Db 1025 CCAAGATGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGT 1084
Qy 1125 TACTATCCAGGCGCACTACCGCCCAACATCACTGCAATGGAATATCAAGGTGCCCAAC 1184
Db 1085 TACTATCCAGGCGCACTACCGCCCAACATCACTGCAATGGAATATCAAGGTGCCCAAC 1144
Qy 1185 AACCGGAACCTGAAAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGT 1244
Db 1145 AACCGGAACCTGAAAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGT 1204
Qy 1245 GGTCTCTCCACCAAGGACTATGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 1304
Db 1205 GGTCTCTCCACCAAGGACTATGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 1264
Qy 1305 CAGTTTGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 1364
Db 1265 CAGTTTGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 1324
Qy 1365 TACACGACACCGGGTCTCTAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 1424
Db 1325 TACACGACACCGGGTCTCTAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 1384
Qy 1425 GGGATGTTTCAATGTGCAAGACTGGACGGTGTGATCCGAAAGGAACTGGCGTGCAGCGGTGG 1484
Db 1385 GGGATGTTTCAATGTGCAAGACTGGACGGTGTGATCCGAAAGGAGTGTGATGAGTGTGATG 1444
```

QY 1485 GCAGACTCCCGGATTATAGTATGAGGTACTGCGCATGCAATGCGCACCCACAGTTC 1544
Db 1445 GCCGACTGCACCGACCAACAGCGATGAGTCACTGAGTTCGACGCGCGCCACAGTTC 1504
QY 1545 ACGTGCAAAAACAGTCTCTGCAAGCCCTCTCTGGTCTGTGACAGTGTCAACAGTGT 1604
Db 1505 ACGTCAAGAACAGTCTCTGCAAGCCCTCTCTGGTCTGTGACAGTGTCAACAGTGT 1564
QY 1605 GGGGACGGAAGTGAACGAGGAGGCTGACAGTCTCTGCTGGGAGTTTCAAGTGTTCAT 1664
Db 1565 GGAGACAAACAGCAGCAGCGAGGAGTGTCTGCGGCCAGACCTTCAGTGTTCAT 1624
QY 1665 GGGAGTGTCTCCCTCAGACGACAGAGTGTATGGGAAGGACAACTGTGGAGATGGGTCT 1724
Db 1625 GGGAGTGCCTCTCGAAAGCCAGCAGTGTGCAATGGGAAGGACGAGTGTGGGACGGGTCC 1684
QY 1725 GACGAGGCTTCATGTGACAGCGTGAATCTGCTCTGCAACCAATATACCTACGCTGC 1784
Db 1685 GACGAGGCTCTGCGCCAGAGTGAACGTGTCTGCTTGTACCAACACACCTACGCTGC 1744
QY 1785 CAAATGGCCTCTGTCTGAGCAAGGCAACCTGAGTGTGATGGAAGACGAGCTGTAGC 1844
Db 1745 CTCATGGCTCTGCTGAGCAAGGCAACCTGAGTGTGAGGGAAGGAGTGTAGC 1804
QY 1845 GATGGCTCGATGAGAAAACCTGTGCTGGCTGCATCTCTTACAAACAGGCTGC 1904
Db 1805 GACGCTCAGATGAGAAAGCTGCGACTGTGGCTGCGCTCATTCACGAGACAGGCTCGT 1864
QY 1905 GTGTTGTGTGACAGAAATGCGGACGAGGCGAGTGGCCCTGCGAGGTGAGCTCCAGCC 1964
Db 1865 GTTGTGGGSCACGGATGCGGATGAGGSCAGTGGCCCTGCGAGGTGAGCTGCATGCT 1924
QY 1965 CTGGGCGAGGCCACTTGTGTGGGCGCTGCTCATCTCTCTGCTGCTGCTCTGCA 2024
Db 1925 CTGGGCGAGGCCACTGCGGTGCTTCCCTCATCTCTCCCACTGCTGCTGCTGCC 1984
QY 2025 GCTCATGCTTTCAGATGACAAAATTTCAAGTACTCAGACTACAGATGTGGACGCC 2084
Db 1985 GCACACTGCTACATCGATGACAGAGGATTCAGTACTCAGACCCCGCAGCTGGACGCC 2044
QY 2085 TTCCTGGTCTGCTGGACAGCAAGCGCAGTGCCTCTGGGCTGCGAGGCTGAGCTC 2144
Db 2045 TTCCTGGCTTGCAGCAGCAGCAGCAGCGCCCTGGGCTGCGAGGCTGAGGCTC 2104
QY 2145 AAAGTATATCAACCCCTTCTCAATGATTTCACTTCGACTATGACATCGCTTG 2204
Db 2105 AAGCGCATCATCTCCACCCCTTCTCAATGACTTCACTTCGACTATGACATCGCTG 2164
QY 2205 CTGGAGCTGGAGAGTGGTGGAGTACAGACCGTGTGCGCCCATCTGCTGCTGAT 2264
Db 2165 CTGGAGCTGGAGAACCCGCGAGAGTACAGCTCCATGTGTGCGGCCCATCTGCTGCGGAC 2224
QY 2265 GCTACCCATGCTTCCCTGCTGGCAAGGCCATCTGGGTCAAGGCTGGGGSCACAAAA 2324
Db 2225 GCCTCCCATGCTTCCCTGCGCGCAAGGCCATCTGGGTCAAGGCTGGGGSCACCCAG 2284
QY 2325 GAGGAGGTACCGAGCGCTGATCTTGCAGAGGCTGAGATCCCGTGTATCAACAGAC 2384
Db 2285 TATGAGGCACTGGCGGCTGATCTCTGCAAAAGGCTGAGATCCCGCTCATCAACAGAC 2344
QY 2385 ACCTGTGAGGACCTCATGCCGACAGATCACCCACGATGATGTGTGGGTTCCTC 2444
Db 2345 ACCTGCGAAGACCTCTGCGCCGACAGATCACCGCGGATGATGTGTGGGTTCCTC 2404
QY 2445 AGTGGGGTGTGAGTCTCTGCGAGGCTGACTCTGTGCGCCCTTGTCAAGCGCGGAGAA 2504
Db 2405 AGCGGCGGTGAGTCTCTGCGAGGCTGATTCGCGGAGACCCCTGTTCAGGCTGAGGCG 2464
QY 2505 GATGGCGAATGTTCCAGGCTGGTGTGTGAGCTGGGTGAGGCTGCGCTCAGAGAAC 2564
Db 2465 GATGGCGGATCTTCCAGGCGGTGTGTGAGCTGGGAGAGCGCTGCGCTCAGAGAAC 2524
QY 2565 AAGCCAGCGTGTACAAAGCTCCCTGTGTGAGTTCGGGACTGGATCAAAAGACACACTGGG 2624

Db 2525 AAGCCAGCGTGTACAAAGCTCCCTCTGTTTGGGACTGGATCAAGAGACACTGGG 2584
QY 2625 GTATAGCAGATGACAGACAGCGACCAAAACACCCACAGGATGCCGACATGACA 2684
Db 2585 GTATAGGCGCGGG---GCCACCAAAATGTGTACACCTGCGGGGCCACCCATCGTCCACC 2641
QY 2685 CCTGGATACAGGAGAGAAACACTGACGACATTTATGTGTGCTCCCGCCCGCCACACA 2744
Db 2642 CCATGTGTCAGC-CCTGCGGCTGGAGCTGGACCGCTGACTGCACGAGCGCCC-CCAGA 2699
QY 2745 ACCCAGACTGTGAAGTCACTCTTAGGACTCAGAGT 2780
Db 2700 ACATACACTGTGAAGTCACTCTCCAGGCTCCAAAT 2735

RESULT 14

US-10-112-221A-1
; Sequence 1, Application US/1011221A
; Publication No. US20030166851A1
; GENERAL INFORMATION:
; APPLICANT: Org, Edgar O.
; APPLICANT: Madison, Edwin
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING A TRANSMEMBRANE SERINE PROTEASE 9,
; FILE OF INVENTION: ENCODED POLYPEPTIDES AND METHODS BASED THEREON
; FILE REFERENCE: 24745-1615
; CURRENT APPLICATION NUMBER: US/10/112,221A
; CURRENT FILING DATE: 2002-03-27
; PRIOR APPLICATION NUMBER: 60/279,228
; PRIOR FILING DATE: 2001-03-27
; PRIOR APPLICATION NUMBER: 60/291,501
; PRIOR FILING DATE: 2001-05-15
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 3147
; TYPE: DNA
; ORGANISM: Homo Sapien
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (23)...(2589)
; OTHER INFORMATION: Nucleotide sequence encoding MTSPI
; PUBLICATION INFORMATION:
; AUTHORS: O'Brien, T.J. and Tanimoto, H.
; DATABASE ACCESSION NUMBER: GenBank #AR081724
; DATABASE ENTRY DATE: 2000-08-31
; PATENT DOCUMENT NUMBER: 5,972,616
; PATENT FILING DATE: 1998-02-20
; PUBLICATION DATE: 1999-10-26
US-10-112-221A-1

Query Match 60.6%; Score 1883.2; DB 13; Length 3147;
Best Local Similarity 81.2%; Pred. No. 0;
Matches 2223; Conservative 0; Mismatches 508; Indels 5; Gaps 3;
QY 45 GATCGGACCGCCAAAACCATGGGTAGCAATCGGGCGCGCAAGCGCGGAGGGGGCTCTCAG 104
Db 5 GAGCGGCTCGGGGTACCATGGGAGCGATCGGGCCCGCAAGGGCGGAGGGGGCCCGAG 64
QY 105 GACTTCGGCGCGGAGCTCAAGTACAATCCCGGCTAGAGAACATGAATGGTTGAGGAG 164
Db 65 GACTTCGGCGCGGAGCTCAAGTACAATCCCGGCTAGAGAACATGAATGGTTGAGGAG 124
QY 165 GGTGTGAGGTTCCTGCTCGCAAAATGCCAAGAAAGTGGAGAGCGAGGCCCGCCAGCGC 224
Db 125 GCGTGGAGTTCCTGCGCAAAATGCCAAGAAAGTGGAGAGCGAGGCCCGCCAGCGC 184
QY 225 TGGGTGCTGCTGGTGGCAGTGTCTGCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 284
Db 185 TGGGTGCTGCTGGCAGCGCTGCTGATCGGCTCTCTTGTCTTGTCTGCGGATCGGCTTC 244
QY 285 CTGGTGGCAGCTTCATNTATCGGAATGTGCGGGTTCAAAAAGTCTTCAATGGCAGTCTG 344

Db 245 CTGGTGTGGCAATTCGACGTACCGGACGCTGGGTGTCAGAAAGGCTTCAATGGCTACATG 304
QY 345 AGGATCAAAATAGATCTTTCTGATGCTATGAGAACTCCACCTCCACAGAGTTTATC 404
Db 305 AGGATCAAAATAGAAATTTGTGATGCCATCGAAGAACTCCAATCCACTGAGTTTGTGA 364
QY 405 AGCTGGCCAGCAGGTGAAGAGGCGCTGAAGCTGTGTAAATGAAGTCCCTGTCTCTG 464
Db 365 AGCTTGGCCAGCAAGGTGAAGAGCAGCGCTGAAGCTGTGTACAGCGGAGTCCCATTCCTG 424
QY 465 GGTCCCTACCAAGAAGTCCGCTGTAACTGCTTCACTGAGGGGAGTGTATGCGCTTAC 524
Db 425 GCGCTTACCAAGAAGTCCGCTGTGACGGCTTTCAGCGAGGGCAGCGTCAATGCGCTTAC 484
QY 525 TACTGGTCAGAGTTTCAGCATCCCCCACCACCTGGCAGAGAGGTTGATCGCGCCATGGCT 584
Db 485 TACTGGTCTGAGTTTCAGCATCCCGCAGCACCTGGTGGAGGAGCGCGGTATGGCC 544
QY 585 GTGAGCCAGTTGTAACTATTGGCACCCCGAGACCGGGCACTGGAATCTCTGTGCTAAACA 644
Db 545 GAGGAGCGCGTAGTCATGCTGCCCGCGGGCGCGCTCCCTGAAAGTCCCTTGTGTCAACC 604
QY 645 TCTGTGTGGCCCTTCCCATTTGACCCAGAAATGCTGCAGAGAGCTTCAGGACAAAGCTGTC 704
Db 605 TCAGTGTGGCTTTCCCCAGCGACTCCAAAACAGTACAGAGGACCCAGGACAAAGCTGTC 664
QY 705 AGTTTGGCCCTGCATGCCCATCGTGTGACAGCTGACACGCTTCACTACCCCTGGCTTCCCC 764
Db 665 AGCTTGGCCCTGCACGCCCGCGGTGTGAGCTGATGCGCTTCAACAAGCGCGCTTCCCT 724
QY 765 AACAGTCCCTACCGCGCATGCCCGCTGCAGTGGGTCCTGGCGGGGAGCGCGAGTCT 824
Db 725 GACAGCCCTTACCCCGCTCATGSCCGCTGCGAGTGGGCCCTGCGGGGGGAGCGCGACTCA 784
QY 825 GTGCTGAGCCTCACTTCCGAAGCTTTGATGTGCTCCCTGTGATGATGACATGGCAGTGC 884
Db 785 GTGCTGAGCCTCACTTCCGAGCTTTGACCTTGGCTCCGAGAGCGCGGAGCGGAC 844
QY 885 GTGCTGACCGGTGATGATGAGCTGAGCCCAATGGAACCCCAACGCTGTGGTGGCGCTGTG 944
Db 845 CTGGTGACGGGTGTAACAACCCCTGAGCCCATGAGGCCCAACGCGCTTGGTGCATGTGT 904
QY 945 GGCACTTCTCACCTCTTACAACTGACTTTCCTCTCTCCAGAGAGTCTTCCCTGTGTC 1004
Db 905 GGCACTTACCTCTCTTACAACTGACTTTCCTCTCTCCAGAGAGTCTTCCCTGTGTC 964
QY 1005 ACGCTGATTAACCAATAGTACCGCGGACATCTTGGCTTTGAGGCCACTTTCTTCCAGCTG 1064
Db 965 AACTGATTAACCAACTGAGCGCGGCACTCCCGCTTTGAGGCCACTTCTTCCAGCTG 1024
QY 1065 CCAAGATGAGCAGCTGTGGCGCTTTTGTGATGACACCCCAAGGAGCAATTTAGCAGCCCC 1124
Db 1025 CCTAGGATGAGCAGCTGTGGAGCGGCTTACGTAAAGCCGAGGGGACATTCACAGCCCC 1084
QY 1125 TACTATCCAGGCACTACCGCCCAACATCACTGACATGGAATATCAAGGTGCCCAAC 1184
Db 1085 TACTATCCAGGCACTACCGCCCAACATCACTGACATGGAATATCAAGGTGCCCAAC 1144
QY 1185 AACCGGAACGTGAAGTGGCTTCAAACTCTTCTATCTGTGGACCCCAACGTATACAGTG 1244
Db 1145 AACAGCATGTGAAGTGAAGTCTCAAAATCTTCTTACTGCTGTGAGCCCGGCGTGCCTGG 1204
QY 1245 GGCTCTTGACCAAGGACTATGTGGAGATCAACCGGAGAGTACTCGCGTGAAGTCC 1304
Db 1205 GGCACTCTGCCCAAGGACTAGTGGAGATCAATGGGAGAAATACTCGGAGAGAGTCC 1264
QY 1305 CAGTTTGTGTGACAGCAACAGCAGCAAGATTAAGTTCACCTTCCATCTGATCACTCG 1364
Db 1265 CAGTTTGTGTGACAGCAACAGCAGCAAGATTAAGTTCACCTTCCATCTGATCACTCG 1324
QY 1365 TACAGGACACCGGGTCTTCTAGCTGAGTACCTCTCTTACGACTCCAAACGACCGGTGCCA 1424
Db 1325 TACAGGACACCGGGTCTTCTAGCTGAATACCTCTCTTACGACTCCAGTCCAGTCCCGG 1384

QY 1425 GGGATGTTTATGTCAAGACTGAGCGGTGCTCCGAGGAACTCGCGTCCGACGGCTGG 1484
Db 1385 GGGAGTTTCACTGTCGCGACCGGGCGGTGTATCCGGAAGGAGCTGCGCTGTGTGCTGG 1444
QY 1485 GCAGACTGCCCGGATTATAGTGTAGAGCGTTTACTTGCAGATGCAATGCCACCCACAGTTC 1544
Db 1445 GCCGACTGCACCGACACAGCGATGAGCTCAATGCAAGTTGCGAGCGCGGACCAAGTTC 1504
QY 1545 ACCTGCAAAAACAGTTTCTGCAAGCCCTCTTCTGGGTCTGTGACAGTGTCAACAGTGT 1604
Db 1505 ACCTGCAAAAACAGTTTCTGCAAGCCCTCTTCTGGGTCTGTGACAGTGTCAACAGTGT 1564
QY 1605 GGGACCGAAAGTACGAGGAGGCTGAGCTGCTCTCTGCGAGTTTCAAGTGTCCCAAT 1664
Db 1565 GGAGCAACAGCGACGAGGAGGCTGAGTTGTCCGCGCCAGACCTTCAGGTGTCCCAAT 1624
QY 1665 GGGAAAGTGTCTCCTCAGAGCCAGAAAGTGAATGGAAGGACAACTGTGGAGATGGGTCT 1724
Db 1625 GGGAAAGTGCCTCTCGAAAAAGCCAGCAGTGCATGGAAGGAGCAGCTGTGGGACGGGTCC 1684
QY 1725 GACGAGCTTTCATGTGACAGCGTGAATGCTCTCTTTCGACCAAAATATACCTACCGTGC 1784
Db 1685 GACGAGCTTCTGCTTTCGACAGGAGGCACTCTGAGTGTGACGGAAGGAGGACTGTAGC 1744
QY 1785 CAAAATGCCCTCTGTCTGAGCAAGGGCAACCTGAGTGTGATGGGAAGGACAACTGTGGAGATGGGTCT 1844
Db 1745 CTCAATGGGTCTGCTTTCGACAGGAGGCACTCTGAGTGTGACGGAAGGAGGACTGTAGC 1804
QY 1845 GATGGCTCCGATGAGAAAACTGTGACTGTGGGTGCGATCTTTTACAAAACAGGCTGC 1904
Db 1805 GACGGCTCAGATGAGAAGGACTGCGACTGTGGGTGCGGTCAATTCAGAGACAGGCTCGT 1864
QY 1905 GTGGTGTGGCGCAATGCGAGCGGCGAGTGGCGCTTGGCAGGTTGAGGCTCCAGCC 1964
Db 1865 GTGTGTGGGGGCACTGATGCGATGCGAGTGAAGGGGAGTGGCCCTTGGCAGGTAAGCTGCATGCT 1924
QY 1965 CTGGGCCAGGGCCACTTGTGTGGGGCTCGCTCATCTCTCTGACTGGCTGGTCTCTGCA 2024
Db 1925 CTGGGCCAGGGCCCACTTCTGGGTGCTTCCCTCATCTCTCCCAACTGGCTGGTCTTGGC 1984
QY 2025 GCTCATTTCTTTCAGGATGACAAAAATTTCAAGTACTCAGACTACAGATGTGGACGGCC 2084
Db 1985 GCACACTGCTACATCGATGACAGAGGATTCAGGTACTCAGACCCCAAGCAGTGCAGCC 2044
QY 2085 TTCTGGGTCTGTGACACAGCAAGCGAGTGGCTTGGGGTGCAGGAGCTGAAGCTC 2144
Db 2045 TTCTGGGTCTGTGACACAGCAAGCGAGCGCGCCCTGGGGTGCAGGAGCGAGGCTC 2104
QY 2145 AAACGTATCATACCCACCCCTTCTTCAATGATTTCACTTTCGACTATGACATCGCCTTG 2204
Db 2105 AAGCGCATCATCTCCACCCCTTCTTCAATGACTTCACCTTTCGACTATGACATCGCGCTG 2164
QY 2205 CTGGAGCTGGAGAGTGTGGTGGAGTACAGCACCGTGTGGCGCCCACTCTGCTGGCTGTAT 2264
Db 2165 CTGGAGCTGGAGAAACCGGACAGTACAGTCTCATGTGGCGGCCCATCTGCTCTGCGGAC 2224
QY 2265 GCTACCCCATCTTCTCTGTGCAAGGCCATCTGGGTACAGGCTGGGGGACACAAAA 2324
Db 2225 GCCTCCCATGTCTTCTCTGCGGCAAGGCCATCTGGGTACGGGCTGGGGACACCCAG 2284
QY 2325 GAGGAGGATACCGAGCGCTGATCTTTCAGAAAGGTTGAGATCCGCTGTCAATCAACAGACC 2384
Db 2285 TATGGAGGCACTGGCGCTGATCTTTCGAAAAAGGTTGAGATCCGCTGTCAATCAACAGACC 2344
QY 2385 ACCTGTGAGACCTCATGCGGACAGATCAACCCACGAATGATGTGTGGGTTCCTC 2444
Db 2345 ACCTGTGAGAAACCTCTTCCGCGACAGATCAACCGCGCATGATGTGTGGGTTCCTC 2404
QY 2445 AGTGGGGTGTGACTTCTGCCAGGGTGAATCTGTGGTGGCCCTTGTCAAGCGGAGAAA 2504
Db 2405 AGCGCGCGGTGGACTCTCCAGGGGTGATTCGGGGGACCCCTGTCTCAGCGTGGAGGG 2464

```
QY 2505 GATGGCGAATGTTCCAGGCTGGTGTGAGAGCTGGGGTGAAGGTGGCTCAGAGGAAC 2564
Db 2465 GATGGCGGATCTTCCAGGCGGTGTGAGTGGGAGAGCGGTGGCTCAGAGGAAC 2524
QY 2565 AAGCCAGGCTGTACACAGAGGCTCCCTGTAGTTCGGGACTGGATCAAGAGCACACTGGG 2624
Db 2525 AAGCCAGGCTGTACACAGAGGCTCCCTGTAGTTCGGGACTGGATCAAGAGAACACTGGG 2584
QY 2625 GTATAGCAGCTGAGCAGACAGCCGACCAACACACACCCACAGGAGTGCCTCCGACATGCACA 2684
Db 2585 GTATAGGCGCGGG--GCCACCAAAATGTGTACACCTTGGCGGGCCACCCATGTCTCCACC 2641
QY 2685 CTTGATACAGAGAGGAACTGAGGACATTTAGTGTGGCTTCCCTCCCTCCCTCCCTCCCTCC 2744
Db 2642 CCAGTGTGCACG-CCTGCGAGGTGGAGACTGGACCGCTGACTGCAACCGGCC-CCAGA 2699
QY 2745 ACCGAGACTGTGAACCTGCATCTTAGGACTCAGAGT 2780
Db 2700 ACATACACTGTGAACCTCAATCTCCAGGGCTCCAAAT 2735

RESULT 15
US-10-112-221A-3
; Sequence 3, Application US/1011221A
; Publication No. US20030166851A1
; GENERAL INFORMATION:
; APPLICANT: Madison, Edwin
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING A TRANSMEMBRANE SERINE PROTEASE 9
; FILE OF INVENTION: ENCODED POLYPEPTIDES AND METHODS BASED THEREON
; FILE REFERENCE: 24745-1615
; CURRENT APPLICATION NUMBER: US/10/112,221A
; PRIOR FILING DATE: 2002-03-27
; PRIOR APPLICATION NUMBER: 60/279,228
; PRIOR FILING DATE: 2001-03-27
; PRIOR APPLICATION NUMBER: 60/291,501
; PRIOR FILING DATE: 2001-05-15
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 3147
; TYPE: DNA
; ORGANISM: Homo Sapien
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1865)...(2590)
; OTHER INFORMATION: Nucleic acid sequence of protease domain of MTP1
US-10-112-221A-3

Query Match 60.6%; Score 1883.2; DB 13; Length 3147;
Best Local Similarity 81.2%; Pred. No. 0;
Matches 2223; Conservative 0; Mismatches 508; Indels 5; Gaps 3;

QY 45 GATCGGACCGCCAAACCATGGGTAGCAATCGGGCGCGCAAGCGCGGAGGGGGCTCTCAG 104
Db 5 GAGCGGCTCGGGGTACCATGAGGAGCGTTCGGCCCGCAAGCGGAGGGGGCCCGAG 64
QY 105 GACTTCGGCGCGGACTCAAGTACAACTCCCGGCTAGAGAAATGAATGGCTTTGAGGAG 164
Db 65 GACTTCGGCGCGGACTCAAGTACAACTCCCGGCTAGAGAAATGAATGGCTTTGAGGAG 124
QY 165 GGTGTGGAGTTCTGCTGCGAACHATGCCAAGAAGTGGAGAAGGAGGAGGAGGAGGAGGAGG 224
Db 125 GGGCGTGGAGTTCTGCTGCGAACHATGCCAAGAAGTGGAGAAGGAGGAGGAGGAGGAGGAGG 184
QY 225 TGGGTGGTGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 284
Db 185 TGGGTGGTGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 244
QY 285 CTGGTGTGGCACTTCCATTAATGAGGAGTGTGGGTTCAAAAAGTCTTCAATGGCCATCTG 344
Db 245 CTGGTGTGGCACTTCCATTAATGAGGAGTGTGGGTTCAAAAAGTCTTCAATGGCTACATG 304
```

```
QY 345 AGGATCAAAATGAGATCTTTTGTGATCGTATGAGAACTCCACCTCCACAGAGTTTATC 404
Db 305 AGGATCAAAATGAGAAATTTTGTGATGCTTACGAGAACTCCAACTCCACTGAGTTGTA 364
QY 405 AGCTTGGCGCAGCAGGCTGAAAGAGCGGCTGAAGCTGTGTAGTGTGTAACAATGAAGTCCCTTCCTG 464
Db 365 AGCTTGGCGCAGCAGGCTGAAAGAGCGGCTGAAGCTGTGTAGTGTGTAACAAGGAGTCCCACTTCCTG 424
QY 465 GGTCTCTTACCAAAAGCTCGGCTGTAACTGCTTCACTGAGGGGAGTGTCTATCCCTTAC 524
Db 425 GGCCTCTACACAGAGGCTCGGCTGTGAGGCTTACGAGGAGGAGGCTGTATCCCTTAC 484
QY 525 TACTGGTCAAGATTCAAGATCCCTCCCAACACTGGCAGAGAGGTTGATCGGCCATGGCT 584
Db 485 TACTGGTCAAGATTCAAGATCCCTCCCAACACTGGTGGAGGAGGAGGCTGTATCCCTTAC 544
QY 585 GTGAGCGGAGTTGTAACATTTGCCACCCCGAGCAGCGGCACTGAAATCTCTTCGTGTACA 644
Db 545 GAGGAGCGGCTAGTCACTGTGCCCCCGCGGCGGCTCCCTGAAAGTCTTTGTGTGTCACC 604
QY 645 TCTGTGTGGCTTCCCTCAATTTGACCCCAAGATGCTGCAGAGGACTCAGACAAACAGCTGC 704
Db 605 TCAGTGGTGGCTTTCCCAACGACTCCAAACAGTACAGAGGAGCCCGAGACAAACAGCTGC 664
QY 705 AGTTTGGCCTGCATGCCCCATGTGTGACAGTGTGACAGCTGTCACTACCCCTGGCTTCC 764
Db 665 AGTTTGGCCTGCATGCCCCATGTGTGAGAGTGTGATGCGCTTCAACACGCGCGCTTCC 724
QY 765 AACAGTCTCCCTCCCGGCGCATGCGCTGCCAGTGGTGTCTGCGGGGAGCGCCACTCT 824
Db 725 GACAGCCCCCTACCCCGCTCATGCGCGCTGCCAGTGGGCGCTGCGGGGGAGCCCGACTCA 784
QY 825 GTGTGAGCCTCACCTTCCGAAGCTTTGATGTGCTCTCTGTGATGAGCATGCGAGTGCAC 884
Db 785 GTGTGAGCCTCACCTTCCGAAGCTTTGATGTGCTCTCTGTGATGAGCATGCGAGTGCAC 844
QY 885 CTGGTACCGTGTATGATGAGTGTGAGCCCGCATGAGAACCCGCTGTGTGGTGGCTGTGT 944
Db 845 CTGGTACCGTGTATGATGAGTGTGAGCCCGCATGAGAACCCGCTGTGTGGTGGCTGTGT 904
QY 945 GGCACCTTCTCACCTCTCAACCTGACTTTTCTCTCTCTCCAGAACTCTTCTCTTCTGTC 1004
Db 905 GGCACCTTCTCACCTCTCAACCTGACTTTTCTCTCTCTCCAGAACTCTTCTCTTCTGTC 964
QY 1005 AGCTGATTAACCAATPACTGACCGCGCATCATCTGCTTTTGGGCGCACTTTCTTCAGCTG 1064
Db 965 ACCTGATTAACCAACACTGAGCGCGCATCCCGCTTTGAGGCGCACTTTCTTCAGCTG 1024
QY 1065 CCCAAGATGAGCAGCTGTGGCGGCTTTTGTAGTGACACCCAGGAGACATTTAGCAGCGCC 1124
Db 1025 CCTAGGATGAGCAGCTGTGGAGCGCGCTTACGTAAAGCCCGAGGAGACATTTCAACAGCGCC 1084
QY 1125 TACTATCCAGCGCACTTACCGCCCAACATCAACTGCATCGAATATCAAGGTGCCCAAC 1184
Db 1085 TACTATCCAGCGCACTTACCGCCCAACATCAACTGCATCGAATATCAAGGTGCCCAAC 1144
QY 1185 AACCGAAAGTGAAGGTGCTTCAAACTCTTCTATCTGTGGTGGACCCCAAGCTACAGTG 1244
Db 1145 AACCGAAAGTGAAGGTGCTTCAAACTCTTCTATCTGTGGTGGACCCCGAGGCTCTGG 1204
QY 1245 GGCTCTGTCACAGAGGACTATGTGAGATCAACCGGAGAGAGTACTGCGGTGAGAGGTCC 1304
Db 1205 GGCACTTCCCCAAGGACTACTGTGAGATCAATGGGAGAAATACTGCGGAGAGAGGTCC 1264
QY 1305 CAGTTTGTGGTGGAGCAGCAACAGCAGCAAGATTAACAGTCCACTTCTTCTGATCACTCG 1364
Db 1265 CAGTTTGTGGTGGAGCAGCAACAGCAGCAAGATTAACAGTCCACTTCTTCTGATCACTCG 1324
QY 1365 TACAGGACACCGGTTCTTAGCTGAGTACCTCTCTACGACTCCAAACAGCCCGTCCCA 1424
Db 1325 TACAGGACACCGGTTCTTAGCTGAGTACCTCTCTACGACTCCAAACAGCCCGTCCCA 1384
QY 1425 GGGATGTTTCATGTGCAAGACTGGAGCGGTGCATCCGAAAGGAACTGCGCTGCGAGCGGTG 1484
```


Db 1385 GGGCAGTTTACGTGCGGACCGGGCGGTGTATCGGAAGAGAGCTGCGCTGTGATGGCTGG 1444
Qy 1485 GCAGACTGCCCGGATTATAGTGTAGTGGGTTACTTCCGATGCAATGCCACCCACAGTTTC 1544
Db 1445 GCCGACTGCCCGGACACAGCGATGAGTCTAACTGCACTTCCGACCGCGGCCACAGTTTC 1504
Qy 1545 ACGTGCAAAAACCAAGTTCTGCAAGCCCTTCTTGGGTCTGTGACAGTGTCAAGCACTGT 1604
Db 1505 ACGTGCAAGAACAAAGTTCTGCAAGCCCTTCTTGGGTCTGCGACAGTGTGAAACGACTGC 1564
Qy 1605 GGGACCGAAGTGTAGGAGGAGGCTGACGTCTCTGCTGGGAGTTTCAAGTGTCCAAAT 1664
Db 1565 GGACAGCAACAGCGACGACGAGCGGTGCAAGTGTCTCGGCCCAAGACCTTCAGGTGTCCAAAT 1624
Qy 1665 GGGAAAGTGTCTCCCTCTCAGAGCGAGAGTGTAAATGGGAAGGACCACTGTGGAGATGGGTCT 1724
Db 1625 GGGAAAGTGTCTCGMAAAGCCAGCAGTGTCAATGGGAAGGACGACTGTGGGAGCGGTTC 1684
Qy 1725 GACGAGGCTTCATGTGACAGCGTGAATGTCTCTTGTGACCAAAATATACCTACCGCTGC 1784
Db 1685 GACGAGGCTCTCTGCCCCCAAGGTGAACGTCTCACTTGTACCAAAACACCACTACCGCTGC 1744
Qy 1785 CAAATGGCTCTGTCTGACGAAGGGCAACCTGAGTGTGATGGGAAGACGGAAGTGTAGC 1844
Db 1745 CTCAATGGCTCTGTCTGACGAAGGGCAACCTGAGTGTGACGGGAAGGAGGACTGTAGC 1804
Qy 1845 GATGGCTCCGATGAGAAAACTGTGACTGTGGGTGCGATCTTTTACCAAAAGGCTCGC 1904
Db 1805 GACGCTCAGATGAGAGGACTGTGACTGTGGGTGCGGTCACTTACGAGACAGGCTCGT 1864
Qy 1905 GTGGTTGGTGCAAGAAATGCGGACGAGGGCGAGTGGCCCTGGCAGGTGAGCTCCACGCC 1964
Db 1865 GTTGTGGGGCACGGATGCGGATGAGGGCGAGTGGCCCTGGCAGGTAAAGCTGCATGCT 1924
Qy 1965 CTGGGCGAGGCGCACTGTGTGGGGCTCGCTCATCTCTCTGACTGGCTGGTCTCTGCA 2024
Db 1925 CTGGGCGAGGCGCACTGTGGGGTCTTCCCTCATCTCTCCAACTGGCTGGTCTCTGCC 1984
Qy 2025 GCTCATTTCTTTCAGGATGACAAAAATTTCAAGTACTCAGACTACAGATGTGGACCGCC 2084
Db 1985 GCACACTGCTACATCGATGACAGAGGATTCAGTACTCAGACCCCAACGAGTGGACCGCC 2044
Qy 2085 TTCTGGGTCTGTGGACCGACGAGCAAGCGAGTGCCTCTGGGGTGCAGGAGCTGAAGCTC 2144
Db 2045 TTCTGGGGTTGCAGACACGAGCAGCGAGCGCCCTGGGGTGCAGGAGCGCAGGCTC 2104
Qy 2145 AAAGTATCATCACCCACCTTCTTCAATGATTTCACTTCCGACTATGACATCGCCTTG 2204
Db 2105 AAGCGCATCATCTCCACCCCTTCTTCAATGACTTCACCTTCGACTATGACATCGCGCTG 2164
Qy 2205 CTGGAGCTGGAGAAAGTGGTGGAGTACAGCACCGTCTGTGGGCCCATCTGCTGCTCTGAT 2264
Db 2165 CTGGAGCTGGAGAAACCGGACAGTACAGCTCCATGTGTGGGCCCATCTGCTGCGCGAC 2224
Qy 2265 GCTACCCATGTCTTCCCTGTGGCAAGGCCATCTGGGTTCACAGCTGGGGGCACACAAA 2324
Db 2225 GCCTCCCATGTCTTCCCTGTGGCAAGGCCATCTGGGTTCACAGCTGGGGGCACACCCAG 2284
Qy 2325 GAGGAGGTACCGGAGCGCTGATCTCTGAGAAAGGTGAGATCGGTGTATCAACACGAGACC 2384
Db 2285 TATGAGGCACTGGCGGTGATCTCTGAAAGGGTGAATCCCGGTCTATCAACGAGACC 2344
Qy 2385 ACCTGTGAGGACCTCATGCCGACGAGATCACCCACGAATGATGTGTGGGTTCCTC 2444
Db 2345 ACCTGGAGAAACCTCTTCCGACGACGATCACGCGCGCATGATGTGGTGGGTTCCTC 2404
Qy 2445 AGTGGGGTGTGGAATCTTCCGACGGTGAATCTGTGGTGGCCCTTGTCAAGCGCGGAGAA 2504
Db 2405 AGCGGCGGCTGGGACTCTCTGACGGGTGATTTCCGGGGGACCCCTGTCCAGCGTGGAGCG 2464
Qy 2505 GATGGGCGAATGTTCCAGGCTGTGTGTGAGTGGGGTGAAGCTCGGCTCAGAGAAC 2564

Db 2465 GATGGGCGGATCTTCCAGGCCGTGTGTGAGCTGGGGAGACGGCTCGCTCAGAGGAAC 2524
Qy 2565 AAGCCAGGCTGTACACAAAGGCTCCCTGTAGTTCCGGGACTGGATCAAAGAGCACACTGGG 2624
Db 2525 AAGCCAGGCTGTACACAAAGGCTCCCTGTGTGGGACTGGATCAAAGAGAACACTGGG 2584
Qy 2625 GTATAGCAGCATGGACAGACAGCGGACACAAACACCCACAGGGATCCCGACATGCACA 2684
Db 2585 GTATAGGGGCGGG---GCCACCAATGTGTACCTGTGGGGCCACCCATCGTCCACC 2641
Qy 2685 CCTGGATACAGGAGAGAAACACTGACGACATTTATGTGTGGCTTCCCCCCCCCAACACA 2744
Db 2642 CCAGTGTGCAGC-CCTGCAAGGCTGGAGACTGGACCGCTGACTGCACCGGCGCC-CCAGA 2699
Qy 2745 ACCCAGACTGTGAAGTGCATCCTTTAGGACTCAGAGT 2780
Db 2700 ACATACACTGTGAACCTCAATCTCCAGGGCTCCAAT 2735

Search completed: February 21, 2004, 10:23:01
Job time : 1070 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 21, 2004, 02:06:37 ; Search time 6211 Seconds
(without alignments)
12154.195 Million cell updates/sec

Title: US-09-900-751-1

Perfect score: 3106

Sequence: 1 catgtagacgctgccccgg.....ttaaaaaaaaaaaaaaa 3106

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

```

1: em_estba.*
2: em_esthum.*
3: em_estin.*
4: em_estmu.*
5: em_estov.*
6: em_estpl.*
7: em_estro.*
8: em_hic.*
9: gb_est1.*
10: gb_est2.*
11: gb_hic.*
12: gb_est3.*
13: gb_est4.*
14: gb_est5.*
15: em_estfun.*
16: em_estom.*
17: em_gss_hum.*
18: em_gss_inv.*
19: em_gss_pln.*
20: em_gssprt.*
21: em_gss_fun.*
22: em_gssmam.*
23: em_gssmus.*
24: em_gsspro.*
25: em_gssrod.*
26: em_gssphg.*
27: em_gssvrl.*
28: gb_gss1.*
29: gb_gss2.*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	3102.8	99.9	3248	11	BC008514	BC008514 Mus muscu
2	3087.6	99.4	4017	11	AK052738	AK052738 Mus muscu
3	956.4	30.8	2689	11	AK085410	AK085410 Mus muscu
4	843.8	27.2	944	13	BQ885551	BQ885551 AGENCOURT

5	833.6	26.8	884	12	BI688367	BI688367 603316692
6	788.8	25.4	803	12	BI690615	BI690615 603314210
7	778.2	25.1	990	13	BQ951529	BQ951529 AGENCOURT
8	760.4	24.5	762	12	BQ919617	BQ919617 602823173
9	759.4	24.4	874	14	CB208872	CB208872 AGENCOURT
10	747.2	24.1	798	12	BI664081	BI664081 603289385
11	739.2	23.8	746	14	BY756896	BY756896 BY756896
12	739.2	23.8	846	12	BI416218	BI416218 602989878
13	735.8	23.7	1116	10	BF301347	BF301347 602029796
14	732	23.6	786	12	BI646380	BI646380 603276244
15	721	23.2	928	12	BQ915045	BQ915045 602814325
16	720.8	23.2	777	12	BI150180	BI150180 602848502
17	716	23.1	928	12	BI694441	BI694441 603348025
18	710.8	22.9	838	12	BI656113	BI656113 603284774
19	707.4	22.8	767	12	BI695744	BI695744 603345094
20	706.8	22.8	840	12	BI453178	BI453178 603173173
21	706.2	22.7	939	14	CA976536	CA976536 AGENCOURT
22	692.6	22.3	725	12	BI248406	BI248406 602967029
23	689.6	22.2	960	12	BI414245	BI414245 602986510
24	687.2	22.1	853	10	BF540669	BF540669 602052757
25	684.4	22.0	697	12	BI853355	BI853355 603379936
26	681.8	22.0	708	12	BI078327	BI078327 602872681
27	681.2	21.9	838	13	BU152438	BU152438 AGENCOURT
28	678.2	21.8	1035	10	BF119461	BF119461 601758388
29	671	21.6	671	12	BG861619	BG861619 602795392
30	665.4	21.4	836	13	BU517436	BU517436 AGENCOURT
31	658.6	21.2	728	12	BI112917	BI112917 602898206
32	650.4	20.9	927	10	BF578780	BF578780 602094810
33	647.4	20.8	649	12	BI697395	BI697395 603348134
34	643.8	20.7	698	14	BY756283	BY756283 BY756283
35	633.6	20.4	886	12	BI657843	BI657843 603284649
36	630.4	20.3	924	12	BI454477	BI454477 603171439
37	623.8	20.1	748	10	BF584458	BF584458 602098352
38	623.4	20.1	854	12	BI416184	BI416184 602989813
39	621.4	20.0	765	10	BF118535	BF118535 601754802
40	617.2	19.9	630	10	BF320582	BF320582 u254902.Y
41	613	19.7	613	14	CA551053	CA551053 C0838D07-
42	612.4	19.7	868	13	BU179919	BU179919 AGENCOURT
43	610.8	19.7	698	12	BG870626	BG870626 602791606
44	610.4	19.7	802	12	BI149332	BI149332 602848871
45	608.8	19.6	790	12	BG968045	BG968045 602832425

ALIGNMENTS

RESULT 1
BC008514
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
REMARK
COMMENT

BC008514 3248 bp mRNA linear HTC 20-SEP-2002
Mus musculus, suppression of tumorigenicity 14 (colon carcinoma),
clone IMAGE:3489914, mRNA.

BC008514
BC008514.1 GI:14250187
HTC.

Mus musculus (house mouse)
Mus musculus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 3248)
Strausberg, R.
Direct Submission
Submitted (25-MAY-2001) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
NIH-MGC Project URL: <http://mgc.nci.nih.gov>
Contact: MGC help desk
Email: cgapbs-remail.nih.gov
Tissue Procurement: Lothar Hennighausen Ph.D., Robin Humphreys
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILNL)
DNA Sequencing by: Sequencing Group at the Stanford Human Genome
Center, Stanford University School of Medicine, Stanford, CA 94305

Web site: <http://www.shgc.stanford.edu>
 Contact: (Dickson, Mark) mcdpaxil.stanford.edu
 Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers,
 R. M.

Clone distribution: MGC clone distribution information can be found
 through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
 Series: IRAP Plate: 7 Row: h Column: 16
 This clone was selected for full length sequencing because it
 passed the following selection criteria: matched mRNA gi: 6677898
 This clone has the following problem: no 5' EST match.

FEATURES

Location/Qualifiers
 1..3248
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="IMAGE:3489914"
 /tissue_type="Mammary tumor. WAP-TGF alpha model. 7 months
 old, gross tissue"
 /clone_lib="NCI CGAP Mam5"
 /lab_host="DH10B"
 /note="Vector: pCMV-SPORT6"
 BASE COUNT 734 a 918 c 928 g 668 t
 ORIGIN

Query Match 99.9%; Score 3102.8; DB 11; Length 3248;
 Best Local Similarity 99.9%; Pred. No. 0;
 Matches 3104; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

QY 1 CATGTTAGACGGCTGCCGGAGGACACACGGCTCTGAGACCGGCTCGGACCGCCAAA 60
DB 111 CATGTTAGACGGCTGCCGGAGGACACACGGCTCTGAGACCGGCTCGGACCGCCAAA 170
QY 61 CCATGGGTAGCAATCGGGGCGCAAGCGCGAGGGGGCTCTCAGGACTTCGGCGGGAC 120
DB 171 CCATGGGTAGCAATCGGGGCGCAAGCGCGAGGGGGCTCTCAGGACTTCGGCGGGAC 230
QY 121 TCAAGTCAACCTCCCGGTAGAGACATGATGCTTGGAGAGGGTGTGAGTTCCTGC 180
DB 231 TCAAGTCAACCTCCCGGTAGAGACATGATGCTTGGAGAGGGTGTGAGTTCCTGC 290
QY 181 CTGGCAACAATGCCAAGAAATGGAGAACGAGGCCCGAGCGCTGGTGTGCTGTGG 240
DB 291 CTGGCAACAATGCCAAGAAATGGAGAACGAGGCCCGAGCGCTGGTGTGCTGTGG 350
QY 241 CAGTGTGTTGAGCTTCTCTTGTCTCCCTCATGCTGGTGGCTTCTGCTGTGGCACTTC 300
DB 351 CAGTGTGTTGAGCTTCTCTTGTCTCCCTCATGCTGGTGGCTTCTGCTGTGGCACTTC 410
QY 301 ATTATCGGAATGTCGGGTTCAAAAGTCTTCAATGGCCATCTGAGGATCAAAATGAGA 360
DB 411 ATTATCGGAATGTCGGGTTCAAAAGTCTTCAATGGCCATCTGAGGATCAAAATGAGA 470
QY 361 TCTTTCTGGATGCGTATGAGAACTCCACCTCCACAGAGTTTATCAGCCTGGCCAGCAGG 420
DB 471 TCTTTCTGGATGCGTATGAGAACTCCACCTCCACAGAGTTTATCAGCCTGGCCAGCAGG 530
QY 421 TGAAGAGGGGCTGAAGCTGTGTACAAATGAAGTCCCTGCTCCCTGGGTCCCTTACCACA 480
DB 531 TGAAGAGGGGCTGAAGCTGTGTACAAATGAAGTCCCTGCTCCCTGGGTCCCTTACCACA 590
QY 481 AGTGGGTGTAACTGCTTCAAGTCAAGGAGGATGATCGGCTACTACTGTGTGAGTTCA 540
DB 591 AGTGGGTGTAACTGCTTCAAGTCAAGGAGGATGATCGGCTACTACTGTGTGAGTTCA 650
QY 541 GCATCCCGCCACCTCGGAGAGAGGTTGATCGGCTGCTGTGAGGAGGAGTTGTA 600
DB 651 GCATCCCGCCACCTCGGAGAGAGGTTGATCGGCTGCTGTGAGGAGGAGTTGTA 710
QY 601 CATTTGCCACCCGAGCAGGGCACTGAATCTCTGCTGTGCTAACTCTGTGGTGGCTTCC 660
DB 711 CATTTGCCACCCGAGCAGGGCACTGAATCTCTGCTGTGCTAACTCTGTGGTGGCTTCC 770

```

```

QY 661 CCATTGACCCAGCAATGCTGCAGAGGACTCAGGACAAACAGCTGCAGTTTGCCTTGCATG 720
DB 771 CCATTGACCCAGCAATGCTGCAGAGGACTCAGGACAAACAGCTGCAGTTTGCCTTGCATG 830
QY 721 CCCATGGTGCAGAGTGCACAGCTTCTACTACCCCTGGCTTCCCAACAGTCCCTACCCGG 780
DB 831 CCCATGGTGCAGAGTGCACAGCTTCTACTACCCCTGGCTTCCCAACAGTCCCTACCCGG 890
QY 781 CGCATGCCGCTGCAGGGGTCTGCGGGGGAGCCGACTCTGTGTGAGCTTCACTTCACT 840
DB 891 CGCATGCCGCTGCAGGGGTCTGCGGGGGAGCCGACTCTGTGTGAGCTTCACTTCACT 950
QY 841 TCCGAAGCTTTGATGCTCGCTCCCTGTGATGAGCATGGCAGTGCCTGTGTGAGCTTCACT 900
DB 951 TCCGAAGCTTTGATGCTCGCTCCCTGTGATGAGCATGGCAGTGCCTGTGTGAGCTTCACT 1010
QY 901 ATAGCTGAGCCCCCATGGAACCCCAACGCTGTGTGCGGCTGTGTGAGCACTTCTCACCT 960
DB 1011 ATAGCTGAGCCCCCATGGAACCCCAACGCTGTGTGCGGCTGTGTGAGCACTTCTCACCT 1070
QY 961 CCTACAACTGACTTCTCTCTCCCGAGACGCTTCTTCTTCCAGTGCCTGTGTGAGCACTT 1020
DB 1071 CCTACAACTGACTTCTCTCTCCCGAGACGCTTCTTCTTCCAGTGCCTGTGTGAGCACTT 1130
QY 1021 CTGACCCGCGACATCTCTGGCTTTGAGGCCACTTCTTCCAGTGCCTGCAGATGAGCAGCT 1080
DB 1131 CTGACCCGCGACATCTCTGGCTTTGAGGCCACTTCTTCCAGTGCCTGCAGATGAGCAGCT 1190
QY 1081 GTGGCGCTTTTGTGAGTGACACCAACGAGCATTTAGAGGCCCTTCTATCCAGGCCACT 1140
DB 1191 GTGGCGCTTTTGTGAGTGACACCAACGAGCATTTAGAGGCCCTTCTATCCAGGCCACT 1250
QY 1141 ACCCGCCCAACATCAACTGCACATGGAATCAAGTGCCTCAACACCGGACCTGAGG 1200
DB 1251 ACCCGCCCAACATCAACTGCACATGGAATCAAGTGCCTCAACACCGGACCTGAGG 1310
QY 1201 TGGCTTCAAACTCTTCTATCTGTGGACCCCAACGCTTACAGTGGCTTCTTGCACCAAG 1260
DB 1311 TGGCTTCAAACTCTTCTATCTGTGGACCCCAACGCTTACAGTGGCTTCTTGCACCAAG 1370
QY 1261 ACTATGTGGAGATCAACGGGGAGAGTACTGCGGTGAGAGGTCCAGTTTGTGTGAGCA 1320
DB 1371 ACTATGTGGAGATCAACGGGGAGAGTACTGCGGTGAGAGGTCCAGTTTGTGTGAGCA 1430
QY 1321 GCAACAGCAGCAAGATTTACAGTCCACTTCCATTTCTGATCACTCGTACACGACACCGG 1380
DB 1431 GCAACAGCAGCAAGATTTACAGTCCACTTCCATTTCTGATCACTCGTACACGACACCG 1490
QY 1381 TCCTAGCTGAGTACCTCTCTAGACTCCAAAGCTCCAAAGCTCCAGGGATTTTCAATGCA 1440
DB 1491 TCCTAGCTGAGTACCTCTCTAGACTCCAAAGCTCCAAAGCTCCAGGGATTTTCAATGCA 1550
QY 1441 AGACTGAGCGTGCATCCGAAAGAACTGCGCTGCGACGCTGCGGAGCTGCGCGATT 1500
DB 1551 AGACTGAGCGTGCATCCGAAAGAACTGCGCTGCGACGCTGCGGAGCTGCGCGATT 1610
QY 1501 ATAGTATGAGCGCTTCTGCGGATGCAATGCCACCAACAGTTCAGTGCAGAAACAGT 1560
DB 1611 ATAGTATGAGCGTGCATCCGAAAGAACTGCGCTGCGACGCTGCGGAGCTGCGCGATT 1670
QY 1561 TCTGCAAGCCCTTCTTCTGCGGTCTGTGACAGTGTCAAGCTGTGGGAGCGGAAGTGA 1620
DB 1671 TCTGCAAGCCCTTCTTCTGCGGTCTGTGACAGTGTCAAGCTGTGGGAGCGGAAGTGA 1730
QY 1621 AGCAGGCTGACGCTCTCTGCTGGGAGTTTCAAGTGTCCATGGAAGTGTCTCCCTC 1680
DB 1731 AGCAGGCTGACGCTCTCTGCTGGGAGTTTCAAGTGTCCATGGAAGTGTCTCCCTC 1790
QY 1681 AGAGCCAGAAAGTGAATGGGAAGGACAACTGTGGAGATGGGTCTGACGAGGCTTCAATG 1740
DB 1791 AGAGCCAGAAAGTGAATGGGAAGGACAACTGTGGAGATGGGTCTGACGAGGCTTCAATG 1850

```


QY 901 ATAGCCTGAGCCCATGGAACCCACGCTGTGGTGGGCTGTGGCACTTCTCACCC 960
Db 1012 ATAGCCTGAGCCCATGGAACCCACGCTGTGGTGGGCTGTGGCACTTCTCACCC 1071
QY 961 CTTACAACTGACTTCTCTCTCCCAAGAGTCTTCTCTGTCAGCTGATTAACCAATA 1020
Db 1072 CTTACAACTGACTTCTCTCTCCCAAGAGTCTTCTCTGTCAGCTGATTAACCAATA 1131
QY 1021 CTGACGGGGAATCTCTGCTTGGAGCCACTTCTTCCAGCTGCCCAGATGAGCAGCT 1080
Db 1132 CTGACGGGGAATCTCTGCTTGGAGCCACTTCTTCCAGCTGCCCAGATGAGCAGCT 1191
QY 1081 GTGGGGCTTTTGTGATGACACCCAGGACATTTAGAGCCCTACTATCCAGGCACT 1140
Db 1192 GTGGGGCTTTTGTGATGACACCCAGGAGCATTTAGAGCCCTACTATCCAGGCACT 1251
QY 1141 ACCCGCCCAACATCACTGACATGGAATATCAAGGTGCCCAACACCGGAACGTGAAGG 1200
Db 1252 ACCCGCCCAACATCACTGACATGGAATATCAAGGTGCCCAACACCGGAACGTGAAGG 1311
QY 1201 TGGCTTCAAACTCTTCTATCTGTGGAGCCCAACAGTACCACTGGGCTCTGACACCAAGG 1260
Db 1312 TGGCTTCAAACTCTTCTATCTGTGGAGCCCAACAGTACCACTGGGCTCTGACACCAAGG 1371
QY 1261 ACTATGTGAGATCAACGGGAGAGTACTGCGGTGAGAGTCCAGTTTGTGGTGAGA 1320
Db 1372 ACTATGTGAGATCAACGGGAGAGTACTGCGGTGAGAGTCCAGTTTGTGGTGAGA 1431
QY 1321 GCAACAGCAGCAAGATTAACAGTCCACTTCTGATCACTCGTACACGGACACCGGGT 1380
Db 1432 GCAACAGCAGCAAGATTAACAGTCCACTTCTGATCACTCGTACACGGACACCGGGT 1491
QY 1381 TCCTAGCTGAGTACTCTCTACAGTCCCAAGTCCAGTCCAGTGGGCTCTGATGTGCA 1440
Db 1492 TCCTAGCTGAGTACTCTCTACAGTCCCAAGTCCAGTCCAGTGGGCTCTGATGTGCA 1551
QY 1441 AGACTGAGCGGTGATCCGAAGGAACTGCGCTGCGAGCGCTGGGCACTGCCGGATT 1500
Db 1552 AGACTGAGCGGTGATCCGAAGGAACTGCGCTGCGAGCGCTGGGCACTGCCGGATT 1611
QY 1501 ATAGTGTAGCGGTACTGCGGATGCAATGCCACCCAGTTCACGTGCAAAACACAGT 1560
Db 1612 ATAGTGTAGCGGTACTGCGGATGCAATGCCACCCAGTTCACGTGCAAAACACAGT 1671
QY 1561 TCTGCAAGCCCTCTTCTGGGTCTGTGACAGTGTCAACGACTGGGGACGGAAGTGAAG 1620
Db 1672 TCTGCAAGCCCTCTTCTGGGTCTGTGACAGTGTCAACGACTGGGGACGGAAGTGAAG 1731
QY 1621 AGGAGGGCTGAGCTGCTCTGCTGGAGTTTCAAGTGTCCAAATGGGAAAGTGTCTCCCTC 1680
Db 1732 AGGAGGGCTGAGCTGCTCTGCTGGAGTTTCAAGTGTCCAAATGGGAAAGTGTCTCCCTC 1791
QY 1681 AGAGCCAGAACTGTAATGGGAAGGACAACTGTGGAGATGGGTCTGACGAGGCTTCATGTG 1740
Db 1792 AGAGCCAGAACTGTAATGGGAAGGACAACTGTGGAGATGGGTCTGACGAGGCTTCATGTG 1851
QY 1741 ACAGCGTAATGTGCTCTCTTGGACCAAAATATACTACCGCTGCCAAAATGGGCTCTGTC 1800
Db 1852 ACAGCGTAATGTGCTCTCTTGGACCAAAATATACTACCGCTGCCAAAATGGGCTCTGTC 1911
QY 1801 TGAGCAAGGGCAACCTGAGTGTGATGGGAGAGGAGTGTAGGATGGCTCCCATGAGA 1860
Db 1912 TGAGCAAGGGCAACCTGAGTGTGATGGGAGAGGAGTGTAGGATGGCTCCCATGAGA 1971
QY 1861 AAAAATGTGACTGTGGGTGCGATCTCTTTTACAAACAGAGCTCGCGTGGTGGTGACGGA 1920
Db 1972 AAAAATGTGACTGTGGGTGCGATCTCTTTTACAAACAGAGCTCGCGTGGTGGTGACGGA 2031
QY 1921 ATGGGAGAGGGGAGTGGCCCTGGGACAGTGAAGCTCCAGCCCTGGGCGACGGGCCACT 1980
Db 2032 ATGGGAGAGGGGAGTGGCCCTGGGACAGTGAAGCTCCAGCCCTGGGCGACGGGCCACT 2091

QY 1981 TGTGTGGGCTCGCTCATCTCTCTGACTGGCTGGTCTCTGAGCTCATTTGCTTTGAG 2040
Db 2092 TGTGTGGGCTCGCTCATCTCTCTGACTGGCTGGTCTCTGAGCTCATTTGCTTTGAG 2151
QY 2041 ATGACAAAAATTTCAAGTACTCAGACTACACGATGTGAGCGGCTTCTGGGTCTGCTGG 2100
Db 2152 ATGACAAAAATTTCAAGTACTCAGACTACACGATGTGAGCGGCTTCTGGGTCTGCTGG 2211
QY 2101 ACCAGAGCAAGCGCAGTGCCTCTGGGGTGACGAGCTGAAGCTCAAACTATCATCACCC 2160
Db 2212 ACCAGAGCAAGCGCAGTGCCTCTGGGGTGACGAGCTGAAGCTCAAACTATCATCACCC 2271
QY 2161 ACCCTTCTTCAATGAATTTCACTTTCGACTATGACATCGCTTGTGGAGCTGGAGAACT 2220
Db 2272 ACCCTTCTTCAATGAATTTCACTTTCGACTATGACATCGCTTGTGGAGCTGGAGAACT 2331
QY 2221 CGGTGAGTACAGCACCTCTGTCGGCCCCCATCTGCTGCTGATGCTACCCATGCTTCC 2280
Db 2332 CGGTGAGTACAGCACCTCTGTCGGCCCCCATCTGCTGCTGATGCTACCCATGCTTCC 2391
QY 2281 CTGCTGGCAAGGCCATCTGGGTTCACAGGCTGGGGSCACACAAAAGAGGAGGTACCGGAG 2340
Db 2392 CTGCTGGCAAGGCCATCTGGGTTCACAGGCTGGGGSCACACAAAAGAGGAGGTACCGGAG 2451
QY 2341 CGCTGATCCTGCAAGAGGTGAGATCCGTGTATCAACACGAGACCACTGTGAGGACCTCA 2400
Db 2452 CGCTGATCCTGCAAGAGGTGAGATCCGTGTATCAACACGAGACCACTGTGAGGACCTCA 2511
QY 2401 TGCCGCGACAGATCACCCACGAAATGATGTGTGGGTTCCTCAGTGGGGGTGTGACT 2460
Db 2512 TGCCGCGACAGATCACCCACGAAATGATGTGTGGGTTCCTCAGTGGGGGTGTGACT 2571
QY 2461 CCTGCCAGGGTGAATCTGTGTGGCCCCCTTGTCAAGCGCGGAGAAAGATGGGCGAAATGTTCC 2520
Db 2572 CCTGCCAGGGTGAATCTGTGTGGCCCCCTTGTCAAGCGCGGAGAAAGATGGGCGAAATGTTCC 2631
QY 2521 AGGTGTGTGTGTGAGCTGGGTGAAAGGTGCGCTCAGAGGAAACAAAGCAGCGGTGTACA 2580
Db 2632 AGGTGTGTGTGTGAGCTGGGTGAAAGGTGCGCTCAGAGGAAACAAAGCAGCGGTGTACA 2691
QY 2581 CAAGGCTCCCTGTAGTTTGGGACTGGATCAAGAGACACACTGGGGTATAGCAGCATGGAC 2640
Db 2692 CAAGGCTCCCTGTAGTTTGGGACTGGATCAAGAGACACACTGGGGTATAGCAGCATGGAC 2751
QY 2641 AGACAGCGGACCAACAAACACCCACAGGGATGCCCGCATGCACACCTGGATACAGGAGAG 2700
Db 2752 AGACAGCGGACCAACAAACACCCACAGGGATGCCCGCATGCACACCTGGATACAGGAGAG 2811
QY 2701 GAACACTGACGACATTTATGCTGTGGCTCCCGCCCCCAACAAACCCAGACTGTGAAT 2760
Db 2812 GAACACTGACGACATTTATGCTGTGGCTCCCGCCCCCAACAAACCCAGACTGTGAAT 2871
QY 2761 GCATCTTTAGGACTCAGAGTTCTTCCAAAGTGGGACCCCTCAAGAGTTGGAGAGAGAC 2820
Db 2872 GCATCTTTAGGACTCAGAGTTCTTCCAAAGTGGGACCCCTCAAGAGTTGGAGAGAGAC 2931
QY 2821 TTGCGTCTAGCGGCCAGCCTCTGGGGCAAGGGTTTGTATGCGAGCCTTCCCTCTAGACC 2880
Db 2932 TTGCGTCTAGCGGCCAGCCTCTGGGGCAAGGGTTTGTATGCGAGCCTTCCCTCTAGACC 2991
QY 2881 CTGAGCTGGGTGAAGATGATGTGTGCCGAGAGCTGCTTCCAACTGTCAATGAGTCTCC 2940
Db 2992 CTGAGCTGGGTGAAGATGATGTGTGCCGAGAGCTGCTTCCAACTGTCAATGAGTCTCC 3051
QY 2941 GCGAGCCCTATGGAGGAGGGGCTCAGGGTCACTCTTTTCCAGGAAGCCAGCCCTAGGA 3000
Db 3052 GCGAGCCCTATGGAGGAGGGGCTCAGGGTCACTCTTTTCCAGGAAGCCAGCCCTAGGA 3111
QY 3001 ACCCCAGAAAGAGTGTGACTTAAGGCTGAAATTTGTTTCTGTGTCAGGGGTGGGTAT 3060
Db 3112 ACCCCAGAAAGAGTGTGACTTAAGGCTGAAATTTGTTTCTGTGTCAGGGGTGGGTAT 3171
QY 3061 TTGAGAGTAAAAACATTTTATTTCTTTTAAAAA 3094

JOURNAL Nature 420, 563-573 (2002)
REFERENCE 6 (bases 1 to 2689)
AUTHORS Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Bono,H., Carninci,P., Fukuchi,J., Aizawa,K., Hanagaki,T., Hara,A., Hashizume,W., Hayashida,K., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T., Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kasukawa,T., Katoh,H., Kawai,J., Kojima,Y., Kondo,S., Konno,H., Kouda,M., Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M., Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Ohnato,N., Okazaki,Y., Saito,R., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N., Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Tagami,M., Tagawa,A., Takahashi,F., Takaku-Akahira,S., Takeda,Y., Tanaka,T., Tomaru,A., Toya,T., Yasunishi,A., Maratsus,M. and Hayashizaki,Y.
TITLE Direct Submission
JOURNAL Submitted (16-APR-2002) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:genome-res@gs.riken.go.jp, URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222, Fax:81-45-503-9216)
COMMENT cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.
Please visit our web site for further details.
URL:http://genome.gsc.riken.go.jp/
URL:http://fantom.gsc.riken.go.jp/
FEATURES
source Location/Qualifiers
1..2689
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="FANTOM_DB:D630022P10"
/db_xref="taxon:10090"
/clone="D630022P10"
/tissue_type="kidney"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="0 day neonate"
1..2689
misc_feature
/notes="suppression of tumorigenicity 14 (colon carcinoma) (MGD|MGRI:1338881, GB|NM_011176, evidence: BLASTN, 100%, match=720)"
BASE COUNT 590 a 747 c 766 g 586 t
ORIGIN
Query Match 30.8%; Score 956.4; DB 11; Length 2689;
Best Local Similarity 76.7%; Pred. No. 1.7e-177;
Matches 1415; Conservative 0; Mismatches 1; Indels 428; Gaps 3;
QY 1 CATGGTAGCGGTGCGCGGAGGACACGCGTCTGAGACCGCGCATCGGACCGCCAAAA 60
DB 110 CATGGTAGCGGTGCGCGGAGGACACCGTCTGAGACCGCGCATCGGACCGCCAAA 169
QY 61 CGATGGGTAGCAATCGGGCCGCAAGCGCGAGAGGGGCTCTCAGGACTTCGGCGCGGGAC 120
DB 170 CGATGGGTAGCAATCGGGCCGCAAGCGCGAGAGGGGCTCTCAGGACTTCGGCGCGGGAC 229
QY 121 TCAAGTACAACCTCCCGGCTAGAGAAATGAATGGCTTTGAGGAGGGGTGTGGAGTCTCTGC 180
DB 230 TCAAGTACAACCTCCCGGCTAGAGAAATGAATGGCTTTGAGGAGGGGTGTGGAGTCTCTGC 289
QY 181 CTGCGAAACAATGCCAAGAAATGGAGAACGAGGCCGCCAGGCGCTCGGTGGTCTCGTGG 240
DB 290 CTGCGAAACAATGCCAAGAAATGGAGAACGAGGCCGCCAGGCGCTCGGTGGTCTCGTGG 349
QY 241 CAGTGTCTGTTCAGTCTCTCTTCTCTCCCTCATGCTCGCTTGGTGTGGCACTTCC 300
DB 350 CAGTGTCTGTTCAGTCTCTCTTCTCTCCCTCATGCTCGCTTGGTGTGGCACTTCC 409
QY 301 ATTATCGGAATGTCGGGTTCAAAAAGTCTTCAATGGCCATCTGAGCATCACAATGAGA 360

Db	410	ATTATCGGAATGTGGGTTTCAAAAAGTCTTCAATGGCCATCTCAGGAATCAAAATGAGA	469
Qy	361	TCCTTCTGGATGCGTATGAGAACTCCACTCCACAGAGTTTATCAG-CTTGCCACAGCCAG	419
Db	470	TCCTTCTGGATGCGTATGAGAACTCCACTCCACAGAGTTTATCAGCCCTGGCCAGCCAG	529
Qy	420	GTGAAGAGGGCGCTGAAGCTGCTGTATCAATGAAGTCCCTGCTGGGTCCTTACCAACAG	479
Db	530	GTGAAGAGGGCGCTGAAGCTGCTGTATCAATGAAGTCCCTGCTGGGTCCTTACCAACAG	589
Qy	480	AAATCGGCTGTAACTGCCCTTCAGTGAAGGCGAGTGTATCGCGCTACTACTGCTCAGAGTTC	539
Db	590	AAATCGGCTGTAACTGCCCTTCAGTGAAGGCGAGTGTATCGCGCTACTACTGCTCAGAGTTC	649
Qy	540	AGCATCCCCCACACTCTGGCAGAGAGAGTTCATCGCGCCATGGCTGTGGAGCGAGTTGTA	599
Db	650	AGCATCCCCCACACTCTGGCAGAGAGAGTTCATCGCGCCATGGCTGTGGAGCGAGTTGTA	709
Qy	600	ACATTGCCACCCGAGACAGGGCTCTGAAATCTCTTGGTGTAAATCTGTGGTGGCCCTTC	659
Db	710	ACATTGCCACCCGAGACAGGGCTCTGAAATCTCTTGGTGTAAATCTGTGGTGGCCCTTC	769
Qy	660	-----	659
Db	770	CGTGAGTGGGGGCAAAACCGAGGTGGGAATGTGTGGGCCAGCCTGTAAAGGGCGGGGCT	829
Qy	660	-----	659
Db	830	GGGAGAGCCACACTCACAGGCCCAGGTGTCTCTGGAGGACTCTGTCTGCTGAGTGGCAGCT	889
Qy	660	-----	659
Db	890	CTTTTGGGTAAATAGGAGGCAAGAGATAAGAAATGATTGCTGGTAAATGTTTGCTT	949
Qy	660	-----CCCATGTGCCCCAGAAATGCTGCAGAGGACTCAGGACA-----	696
Db	950	TCATGTTCTCCACCCATTGACCCAGAAATGCTGCAGAGGACTCAGGACAGTAAGTATCAT	1009
Qy	697	-----	696
Db	1010	GSTGACCTCTGTTGGAGAAAGTGTCTTGCCCTTTAAGCCATGGGTGCCCTTAAAGT	1069
Qy	697	-----	696
Db	1070	CAGATGGTACTGATGGACAAGCATCTCTTCTTTTCCCTTCCGGTCCAAAGCTCCTTTCC	1129
Qy	697	-----	696
Db	1130	TTGTGTCTACTGAGTTGCGAGTGATCTGAGACGTAATTGGGGGTAGAGCTAGCGCGGGG	1189
Qy	697	-----ACAGCTGCGAGTTTTGC-----	712
Db	1190	CGGGGGCTTCAGGGTTTCAGCCTTCTCTTCCATCTCTCCACAGACAGCTGCGAGTTTTGC	1249
Qy	713	CTGTATGCCATGGTGCAGAGTGACAGCTTCACTACCCCTGGCTTCCCCAACAGTCC	772
Db	1250	CTGTGATGCCATGGTGCAGAGTGACAGCTTCACTACCCCTGGCTTCCCCAACAGTCC	1309
Qy	773	CTACCCGGCGCATGCCCGCTGCCAGTGGGTCTGTGGGGGGAGCGCGACTCTGTGCTGAG	832
Db	1310	CTACCCGGCGCATGCCCGCTGCCAGTGGGTCTGTGGGGGGAGCGCGACTCTGTGCTGAG	1369
Qy	833	CTCACCTTCCGAAGCTTTGATGTGCTCTCTGTGATGAGCATGGCAGTGCACCTGGTTCAC	892
Db	1370	CTCACCTTCCGAAGCTTTGATGTGCTCTCTGTGATGAGCATGGCAGTGCACCTGGTTCAC	1429
Qy	893	CGTGTATGATAGCCCTGAGCCCCATGGAAACCCACGCTGTGGTGGGCTGTGTGGCACCTT	952
Db	1430	CGTGTATGATAGCCCTGAGCCCCATGGAAACCCACGCTGTGGTGGGCTGTGTGGCACCTT	1489
Qy	953	CTCACCTTCTCAACCTTGACTTTCTCTCTCTCCAGAAAGTCTCTCTGTCAGCTGAT	1012

Db	1490	CTACCCCTCTA	CA	ACCTGACTTTCTCTCTCTCCAGAA	CGTCTTCTTGTCAGCTGAT	1549
Qy	1013	AACCAATACT	GACCGGCGACATCGTGGCTTTGAGGCCACTTCTTCCAGCTGCCCAAGAT	1072		
Db	1550	AACCAATACT	GACCGGCGACATCTGSGCTTTGAGGCCACTTCTTCCAGCTGCCCAAGAT	1609		
Qy	1073	GAGCAGCTG	GSGGCGCTTTTGTAGTGACACCCAAAGGGA	CA	TTTGTAGCAGCCCTTACTATCC	1132
Db	1610	GAGCAGCTG	TGSGGCGCTTTTGTAGTGACACCCAAAGGGA	CA	TTTGTAGCAGCCCTTACTATCC	1669
Qy	1133	AGGCCACT	ACCCGCCAAACATCAACTGCATCGAATATCAAGCGTGCCCAACACCGGA	1192		
Db	1670	AGGCCACT	ACCCGCCAAACATCAACTGCATCGAATATCAAGCGTGCCCAACACCGGA	1729		
Qy	1193	CGTGAAGGT	CGCGTTCAAACCTTCTATCTGTGGACCCCAACGTCACCATGGGCTCCCTG	1252		
Db	1730	CGTGAAGGT	CGCGTTCAAACCTTCTATCTGTGGACCCCAACGTCACCATGGGCTCCCTG	1789		
Qy	1253	CACCAAGGACT	ATGTGGAGATCAACGGGGAAGTACTGCGGTGAGAGGTCCCAAGTTTGT	1312		
Db	1790	CACCAAGGACT	ATGTGGAGATCAACGGGGAAGTACTGCGGTGAGAGGTCCCAAGTTTGT	1849		
Qy	1313	GGTGAGCAG	CAACAGCAGCAAGATTACGTGCCACTTCCATTCTGTGACTACTCGTACACGGA	1372		
Db	1850	GGTGAGCAG	CAACAGCAGCAAGATTACGTGCCACTTCCATTCTGTGACTACTCGTACACGGA	1909		
Qy	1373	CACCGGGTCT	TAGCTGAGTACCTCTCTCTTAAGACTCCAAAGACC	1416		
Db	1910	CACCGGGTCT	TAGCTGAGTACCTCTCTCTTAAGACTCCAAAGACC	1953		

RESULT 4	ACCESSION	REFERENCE
BQ885551	VERSION	AUTHORS
LOCUS	KEYWORDS	TITLE
DEFINITION	SOURCE	JOURNAL
	ORGANISM	COMMENT

BQ85551 944 bp mRNA linear EST 16-AUG-2002
 AGENCOURT_8748027 NIH_MGC_130 Mus musculus cdna clone IMAGE:6335237
 5' mRNA sequence.

BQ885551
BQ885551.1 GI:22277569
EST.

Mus musculus (house mouse)

Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Mus;
1 (bases 1 to 944)
NIH-MGC <http://mgc.nci.nih.gov/>.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Mark Maconochie, Ph.D. and Nancy L. Freeman,
Ph.D.

CDNA Library Preparation: ResGen, Invitrogen Corp
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: LLAM13797 row: f column: 06
High quality sequence start: 623.

FEATURES	SOURCE
1. Feature 1	Source 1
2. Feature 2	Source 2
3. Feature 3	Source 3
4. Feature 4	Source 4
5. Feature 5	Source 5
6. Feature 6	Source 6
7. Feature 7	Source 7
8. Feature 8	Source 8
9. Feature 9	Source 9
10. Feature 10	Source 10
11. Feature 11	Source 11
12. Feature 12	Source 12
13. Feature 13	Source 13
14. Feature 14	Source 14
15. Feature 15	Source 15
16. Feature 16	Source 16
17. Feature 17	Source 17
18. Feature 18	Source 18
19. Feature 19	Source 19
20. Feature 20	Source 20
21. Feature 21	Source 21
22. Feature 22	Source 22
23. Feature 23	Source 23
24. Feature 24	Source 24
25. Feature 25	Source 25
26. Feature 26	Source 26
27. Feature 27	Source 27
28. Feature 28	Source 28
29. Feature 29	Source 29
30. Feature 30	Source 30
31. Feature 31	Source 31
32. Feature 32	Source 32
33. Feature 33	Source 33
34. Feature 34	Source 34
35. Feature 35	Source 35
36. Feature 36	Source 36
37. Feature 37	Source 37
38. Feature 38	Source 38
39. Feature 39	Source 39
40. Feature 40	Source 40
41. Feature 41	Source 41
42. Feature 42	Source 42
43. Feature 43	Source 43
44. Feature 44	Source 44
45. Feature 45	Source 45
46. Feature 46	Source 46
47. Feature 47	Source 47
48. Feature 48	Source 48
49. Feature 49	Source 49
50. Feature 50	Source 50
51. Feature 51	Source 51
52. Feature 52	Source 52
53. Feature 53	Source 53
54. Feature 54	Source 54
55. Feature 55	Source 55
56. Feature 56	Source 56
57. Feature 57	Source 57
58. Feature 58	Source 58
59. Feature 59	Source 59
60. Feature 60	Source 60
61. Feature 61	Source 61
62. Feature 62	Source 62
63. Feature 63	Source 63
64. Feature 64	Source 64
65. Feature 65	Source 65
66. Feature 66	Source 66
67. Feature 67	Source 67
68. Feature 68	Source 68
69. Feature 69	Source 69
70. Feature 70	Source 70
71. Feature 71	Source 71
72. Feature 72	Source 72
73. Feature 73	Source 73
74. Feature 74	Source 74
75. Feature 75	Source 75
76. Feature 76	Source 76
77. Feature 77	Source 77
78. Feature 78	Source 78
79. Feature 79	Source 79
80. Feature 80	Source 80
81. Feature 81	Source 81
82. Feature 82	Source 82
83. Feature 83	Source 83
84. Feature 84	Source 84
85. Feature 85	Source 85
86. Feature 86	Source 86
87. Feature 87	Source 87
88. Feature 88	Source 88
89. Feature 89	Source 89
90. Feature 90	Source 90
91. Feature 91	Source 91
92. Feature 92	Source 92
93. Feature 93	Source 93
94. Feature 94	Source 94
95. Feature 95	Source 95
96. Feature 96	Source 96
97. Feature 97	Source 97
98. Feature 98	Source 98
99. Feature 99	Source 99
100. Feature 100	Source 100

```

1. .944
   /organism="Mus musculus"
   /mol_type="mRNA"
   /db_xref="taxon:10090"
   /clone="IMAGE:6335237"
   /lab_host="DH10B (phage-resistant)"
   /clone_lib="NIH MGC 130"
   /note="Organ. ofocytes; Vector: pCMV
   Site 1: EcoRV; Site 2: NotI; cloned
   Primer: Oligo dt. Average insert s
   Constructed by ResGen, Invitrogen C
   NIH MGC Library."

```

BASE COUNT	184 a	288 c	272 q	200 t
------------	-------	-------	-------	-------

ORIGIN

Query Match 27.2%; Score 843.8; DB 13; Length 944;
Best Local Similarity 96.3%; Pred. No. 2,7e-155;
Matches 885; Conservative 0; Mismatches 32; Indels 2; Gaps 2;
QY 30 GCGTCTGAGACCGGATCGGACCGCCAAACCATGCGGTAGCAATCGGGCGCGCAAGGCC 89
DB 1 GCGTCTGAGACCGGCGATCGGACCGCCAAACCATGCGGTAGCAATCGGGCGCGCAAGGCC 60
QY 90 GGAGGGGGCTCTCAGGACTTCGGCGGGGACTCAAGTCAACTCCCGCTAGAGCAATG 149
DB 61 GGAGGGGGCTCTCAGGACTTCGGCGGGGACTCAAGTCAACTCCCGCTAGAGCAATG 120
QY 150 AATGCTTTGAGAGGGGTGAGTTCCTGCGAAACAATGCGCAAGAAAGTGGAGAG 209
DB 121 AATGCTTTGAGAGGGGTGAGTTCCTGCGAAACAATGCGCAAGAAAGTGGAGAG 180
QY 210 CGAGGCCCGGCGCTGGGTGCTGCTGGGAGTGGTTCAGCTTCTCTTGTCTCTCC 269
DB 181 CGAGGCCCGGCGCTGGGTGCTGCTGGGAGTGGTTCAGCTTCTCTTGTCTCTCC 240
QY 270 CTCATGCTGGTCTGCTGGTGGCACTTCATATCGGAATGTCGGGTTCAAAAAGTC 329
DB 241 CTCATGCTGGTCTGCTGGTGGCACTTCATATCGGAATGTCGGGTTCAAAAAGTC 300
QY 330 TTCATGCGCCATCTGAGGATCAAAATGAGATCTTTTGGATGGGTATGAGAACTCCACC 389
DB 301 TTCATGCGCCATCTGAGGATCAAAATGAGATCTTTTGGATGGGTATGAGAACTCCACC 360
QY 390 TCCACAGAGTTTATCAGCTGCGCCAGCAGAGTGAAGGAGGCGCTGAAGCTCTGTACAT 449
DB 361 TCCACAGAGTTTATCAGCTGCGCCAGCAGAGTGAAGGAGGCGCTGAAGCTCTGTACAT 420
QY 450 GAAGTCCCTGCTCTGGTCCCTTACCAAGAAGTGGCTGTAACTGCTTCAGTGAAGGC 509
DB 421 GAAGTCCCTGCTCTGGTCCCTTACCAAGAAGTGGCTGTAACTGCTTCAGTGAAGGC 480
QY 510 AGTCTCATCGGCTACTACTGTGAGTTCAGATTCAGATTCCTCCCACTGGCAGAGGTT 569
DB 481 AGTCTCATCGGCTACTACTGTGAGTTCAGATTCAGATTCCTCCCACTGGCAGAGGTT 540
QY 570 GATCGCGCATGCTGTGGAGCGAGTTGTAACTTGCACATTCGCCACCCGAGCAGCGGCTGAAA 629
DB 541 GATCGCGCATGCTGTGGAGCGAGTTGTAACTTGCACATTCGCCACCCGAGCAGCGGCTGAAA 600
QY 630 TCCCTGTGCTTAACTGTGGTGGCTTCCCATTTGACCCCAAGATGCTGCAGGACT 689
DB 601 TCCCTGTGCTTAACTGTGGTGGCTTCCCATTTGACCCCAAGATGCTGCAGGACT 660
QY 690 CAGGACACAGCTGAGTTCCTGCTGATGCTGATGCTGATGCTGATGCTGATGCTGAT 749
DB 661 CAGGACACAGCTGAGTTCCTGCTGATGCTGATGCTGATGCTGATGCTGATGCTGAT 720
QY 750 ACCCTGCTTCCCAACAGTCCCTTACCCGCGCATGCGGCTGCCAGTGGTCTCTCGG 809
DB 721 ACCCTGCTTCCCAACAGTCCCTTACCCGCGCATGCGGCTGCCAGTGGTCTCTCGG 780
QY 810 GGGAGCGCGACTCTGTGCTGAGCTCACTTCCGAAAGCTTTGATGCTGCTTCCCTGTGAT 869
DB 781 GGGAGCGCGACTCTGTGCTGAGCTCACTTCCGAAAGCTTTGATGCTGCTTCCCTGTGAT 840
QY 870 GAGCAT-GGCAAGTACCTGGTCAAGGT-GTATGATAGCTGAGGCCCATGGAACCCCAAG 927
DB 841 GACCATGGCAATTCCTGGCCCCCGGTGATGATTAACCTTGAGCCCCCATGGGACCCCA 900
QY 928 CTGTGGTGGGCTGTGG 946
DB 901 CCCTGGTGGTGGGCTTG 919

RESULT 5
P1688367

LOCUS
DEFINITION 603316692F1 NCI_CGAP_Mam6 884 bp mRNA linear EST 18-SEP-2001
ACCESSION mRNA sequence.
VERSION BI688367
KEYWORDS BI688367.1 GI:15650996
SOURCE EST.
ORGANISM Mus musculus (house mouse)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
NIH-MGC http://mgc.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Jeffrey Green M.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L1AM1906 row: h column: 24
High quality sequence stop: 831.
FEATURES
Location/Qualifiers
1..884
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:5356487"
/sex="female, virgin"
/tissue_type="infiltrating ductal carcinoma"
/dev_stage="5 months"
/lab_host="DH10B"
/clone_lib="NCI_CGAP Mam6"
/notes="Organ: mammary; Vector: pCMV-SPORT6; Site 1: Sali;
Site 2: NotI; Cloned unidirectionally. Primer: Oligo dt.
Library constructed by Life Technologies. Investigator
providing samples: Jeffrey Green, M.D., NIH"

BASE COUNT 197 a 232 c 268 g 187 t
ORIGIN
Query Match 26.8%; Score 833.6; DB 12; Length 884;
Best Local Similarity 98.6%; Pred. No. 2,7e-153;
Matches 872; Conservative 0; Mismatches 9; Indels 3; Gaps 3;
QY 1165 GGAATATCAAGGTGCCCAACAACCGGAACTGAGGTGCGCTTCAAACCTCTTCTATCTGG 1224
DB 1 GGAATATCAAGGTGCCCAACAACCGGAACTGAGGTGCGCTTCAAACCTCTTCTATCTGG 60
QY 1225 TGACACCCCAACGTACAGTGGCTCTGACCAAGGACTATGTGGAGATCAACGGGAGA 1284
DB 61 TGACACCCCAACGTACAGTGGCTCTGACCAAGGACTATGTGGAGATCAACGGGAGA 120
QY 1285 AGTACTCGGTGAGAGGTCCCACTTGTGTGTGAGCAGCAACAGCAGCAAGATTACAGTCC 1344
DB 121 AGTACTCGGTGAGAGGTCCCACTTGTGTGTGAGCAGCAACAGCAGCAAGATTACAGTCC 180
QY 1345 ACTTCATCTTGATCACTCGTACACGGACACCGGGTTCCTAGTGTAGTACCTCTCTACG 1404
DB 181 ACTTCATCTTGATCACTCGTACACGGACACCGGGTTCCTAGTGTAGTACCTCTCTACG 240
QY 1405 ACTCCCAACGACCGCTGCCAGGATCTTCATGTGCAAGACTGACCGGTGCATCCGAAGG 1464
DB 241 ACTCCCAACGACCGCTGCCAGGATCTTCATGTGCAAGACTGACCGGTGCATCCGAAGG 300
QY 1465 AACTGGCTGCGACCGCTGGCAGACTGCGCCGAGATTATAGTGTAGCGGTACTTCCCGAT 1524
DB 301 AACTGGCTGCGACCGCTGGCAGACTGCGCCGAGATTATAGTGTAGCGGTACTTCCCGAT 360
QY 1525 GCAATGCCACCCACCAAGTTCAGTGCAGAAACCAAGTCTTCTGCAAGCCCTCTTCTGCGTCT 1584

DEFINITION AGENCOURT_8952916 NCI_CGAP_Co24 Mus musculus cDNA clone
 IMAGE:6475074 5', mRNA sequence.
 ACCESSION BQ951529
 VERSION BQ951529.1 GI:22367007
 KEYWORDS EST.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE 1 (bases 1 to 990)
 NIH-MGC <http://mgi.nci.nih.gov/>.
 AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished
 COMMENT Contact: Robert Strausberg, Ph.D.
 Tissue Procurement: The Cepko Laboratory
 cDNA Library Preparation: Life Technologies, Inc.
 DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone Distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
 Plate: LLAM14013 row: d column: 19
 High quality sequence stop: 635.
 Location/Qualifiers
 1..990
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="FVB/N"
 /db_xref="taxon:10090"
 /clone="IMAGE:6475074"
 /lab_host="DH10B (TI phase-resistant)"
 /clone_libs="NCI_CGAP_Co24"
 /note="Organ: colon; Vector: pCMV-Sport6; Site 1: NotI;
 Site 2: SalI; Cloned unidirectionally. Primer: Oligo dt.
 Average insert size 1.6 kb. Constructed by Life
 Technologies. Note: this is a NCI_CGAP Library."
 BASE COUNT 229 a 272 c 286 g 201 t 2 others
 ORIGIN
 Query Match 25.1%; Score 778.2; DB 13; Length 990;
 Best Local Similarity 92.9%; Pred. No. 1.9e-142;
 Matches 924; Conservative 0; Mismatches 60; Indels 11; Gaps 10;
 QY 981 TCCTCCAGACGTCCTTCCTGTCAGCTGTATACCAATATGACCGGCGACATCTCTGGC 1040
 Db 1 TCCTCCAGACGTCCTTCCTGTCAGCTGTATACCAATATGACCGGCGACATCTCTGGC 60
 QY 1041 TTTGAGGCGCACTTCTTCCAGCTGCCAAGATGACGAGCTGTGGCGGCTTTTTCAGTGC 1100
 Db 61 TTTGAGGCGCACTTCTTCCAGCTGCCAAGATGACGAGCTGTGGCGGCTTTTTCAGTGC 120
 QY 1101 ACCCAAGGACATTTAGCAGCCCTTACTATCCAGGCCACTACCGGCCCAACATCACTGC 1160
 Db 121 ACCCAAGGACATTTAGCAGCCCTTACTATCCAGGCCACTACCGGCCCAACATCACTGC 180
 QY 1161 ACATGGATATCAGGTGCCCAACACCGNAACGTGAAGTGCCTCAACTCTTCTAT 1220
 Db 181 ACATGGATATCAGGTGCCCAACACCGNAACGTGAAGTGCCTCAACTCTTCTAT 240
 QY 1221 CTGCTGGACCCCAACGTACAGTGGGTCTCTGACCAAGGACTATGTGGAGATCAACGGG 1280
 Db 241 CTGCTGGACCCCAACGTACAGTGGGTCTCTGACCAAGGACTATGTGGAGATCAACGGG 300
 QY 1281 GAGAAGTACTCGGTGAGAGTCCAGTTTGTGTGAGCAGCAGCAGCAAGATTACA 1340
 Db 301 GAGAAGTACTCGGTGAGAGTCCAGTTTGTGTGAGCAGCAGCAGCAAGATTACA 360
 QY 1341 GTCCACTTCCATTTCTGATCACTCTGACGAGACCGGGTTCCTAGCTGAGTACTCTCC 1400
 Db 361 GTCCACTTCCATTTCTGATCACTCTGACGAGACCGGGTTCCTAGCTGAGTACTCTCC 420
 QY 1401 TACGACTCCAAACGACCGGTGCCAGGATGTTTCATGTGCAAGACTGCGACGGTGCATCCGA 1460

Db 421 TACGACTCCAAACGACCGGTGCCAGGATGTTTCATGTGCAAGACTGCGACGGTGCATCCGA 480
 QY 1461 AAGGAATCGCTCGGACGGCTGGGCGACAGTCCCGGATATATAGTATGACGTTACTGCG 1520
 Db 481 AAGGAATCGCTCGGACGGCTGGGCGACAGTCCCGGATATATAGTATGACGTTACTGCG 540
 QY 1521 CGATGCAATGCCACCCACCACTTCACGTTCGCAAAACCAAGTTCGCAAGCCCT-CTTCTG 1579
 Db 541 CGATGCAATGCCACCCACCACTTCACGTTCGCAAAACCAAGTTCGCAAGCCCTCTTCTG 600
 QY 1580 GGTCTGTGACAGTGTCAACGACT-GTGGGACGCAAGTGCAGGAGGGCTGCAGCTGTC 1638
 Db 601 GGTCTGTGACAGTGTCAACGACTGTTGGGACGCAAGTGCAGGAGGGCTGCAGCTGTC 660
 QY 1639 CTGCTGGAGTTTCAAGTGTCCAAATGGGAAGTGTCTCCCTCAGAGCAGAGT-GTAAT 1697
 Db 661 CTGCTGGAGTTTCAAGTGTCCAAATGGGAAGTGTCTCCCTCAGAGCAGAGTGTGTAAT 720
 QY 1698 GGAAGGACCAACTGT--GGAGATGGGTCTGACGAGGTTCATGTG-ACAGCGTGAATGTC 1754
 Db 721 GGAAGGACNCTGTTGGGAGATGGGTCTGACGAGGTTCATGTGACAGCGTGAATGTC 780
 QY 1755 GTCTTTTGACCAATATACCTACCGCTGCCAAATGGCTCTCTGAGCAAGGGCAAC 1814
 Db 781 GTCTTTTGACCA-ATATATACCTACCGCTGCCAAATGG-CTCTGTCTGAGC-AGGGCACC 837
 QY 1815 CCTGAGTGTGATGGAGACGGACTGTAGGATGGCTCCGATGAGAAACTGTGACTGT 1874
 Db 838 CTGATTTGGATGGGAAGCGGACTGTACCGATGG-TCCGATGAAAAAAC-GGGAATGT 895
 QY 1875 GGGTGCAGTCTTTTACCAACAGGCTCGGTGTGTTGTGTCACGAATGGGACGAGGGC 1934
 Db 896 GGGCTGGACCCCTTACCAAGGCTTCCGGTGTGGTGGGAGCAGCAGGAGGG 955
 QY 1935 GAGTGGCCCTGGCAGGTGAGCTCCACGCCCTGGG 1969
 Db 956 CAGGGGCCCGTGGAGGGGACCTCCCGGCGGTGG 990
 RESULT 8
 BQ919617
 LOCUS 602823173F1 NCI_CGAP_Mam6 Mus musculus cDNA clone IMAGE:4951962 5',
 mRNA sequence.
 BQ919617
 ACCESSION BQ919617.1 GI:14300093
 VERSION BQ919617
 KEYWORDS EST.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE 1 (bases 1 to 762)
 NIH-MGC <http://mgi.nci.nih.gov/>.
 AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgabs-remail.nih.gov
 Tissue Procurement: Jeffrey Green M.D.
 cDNA Library Preparation: Life Technologies, Inc.
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone Distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
 Plate: LLAM10909 row: a column: 19
 High quality sequence stop: 727.
 Location/Qualifiers
 1..762
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="FVB/N"
 /db_xref="taxon:10090"
 FEATURES
 source

```
/clone="IMAGE:4951962"
/sex="female, virgin"
/tissue type="infiltrating ductal carcinoma"
/lab stage="5 months"
/lab host="DH10B"
/clone_lib="NCI_CGAP_Mam6"
/note="Organ: mammary; Vector: pCMV-SPORT6; Site 1: SalI;
Site 2: NotI; Cloned unidirectionally. Primer: Oligo dt.
Library constructed by Life Technologies. Investigator
providing samples: Jeffrey Green, M.D., NIH"
BASE COUNT      164 a 243 c 190 g 165 t
ORIGIN
Query Match      24.5%; Score 760.4; DB 12; Length 762;
Best Local Similarity 99.9%; Pred. No. 6.4e-139;
Matches 761; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 682 AGAGGACTCAGACACAGCTGCAGTTTGGCCCTGCATGCCCATGGTCAGCAGTGACAC 741
Db 1 AGAGGACTCAGACACAGCTGCAGTTTGGCCCTGCATGCCCATGGTCAGCAGTGACAC 60
QY 742 GCTTCACTACCCCTGGCTTCCCAACAGTCCCTACCCGGCGCATCCCGCTGCCAGTGGG 801
Db 61 GCTTCACTACCCCTGGCTTCCCAACAGTCCCTACCCGGCGCATCCCGCTGCCAGTGGG 120
QY 802 TCCTCGGGGGGACCCCGACTCTGTGTGAGCCTCACCTTCCGAAGCTTGTATGTCGCTC 861
Db 121 TCCTCGGGGGGACCCCGACTCTGTGTGAGCCTCACCTTCCGAAGCTTGTATGTCGCTC 180
QY 862 CTTGTGATGAGCATGGCAGTGCAGTGTGCTGACCTGACCGTGTATGATAGCCTTCCGAAGCTTGTATGTCGCTC 921
Db 181 CTTGTGATGAGCATGGCAGTGCAGTGTGCTGACCTGACCGTGTATGATAGCCTTCCGAAGCTTGTATGTCGCTC 240
QY 922 CCCAGCTGTGTGGCGGTGTGTGGCACTTCTCACCTTCTTACACCTGACTTTTCCTCT 981
Db 241 CCCAGCTGTGTGGCGGTGTGTGGCACTTCTCACCTTCTTACACCTGACTTTTCCTCT 300
QY 982 CTCCCGAAGCTCTTCTTGTGACGCTGATTAACCAATCTAGTACCGGCGCATCTCGCT 1041
Db 301 CTCCCGAAGCTCTTCTTGTGACGCTGATTAACCAATCTAGTACCGGCGCATCTCGCT 360
QY 1042 TTGAGGCCACTTTCTTCCAGCTGCCCAAGATGAGCAGCTGTGGCGGCTTTTGTAGTGACA 1101
Db 361 TTGAGGCCACTTTCTTCCAGCTGCCCAAGATGAGCAGCTGTGGCGGCTTTTGTAGTGACA 420
QY 1102 CCCAAGGACATTTAGCAGCCCTACTATCCAGGCCACTTACCCGCCCAACATCAACTGCA 1161
Db 421 CCCAAGGACATTTAGCAGCCCTACTATCCAGGCCACTTACCCGCCCAACATCAACTGCA 480
QY 1162 CATGGAATATCAAGTGGCCCAACACCGGAAGCTGAAGTGGGCTTCAAACTTTCTATC 1221
Db 481 CATGGAATATCAAGTGGCCCAACACCGGAAGCTGAAGTGGGCTTCAAACTTTCTATC 540
QY 1222 TGGTGACCCCAACTTACAGTGGGCTCTCTGACCAAGGACTATGTGAGATCAACGGGG 1281
Db 541 TGGTGACCCCAACTTACAGTGGGCTCTCTGACCAAGGACTATGTGAGATCAACGGGG 600
QY 1282 AGAAGTACTGCGGTGAGAGGTCCCAAGTTGTGGTGAGCAGCAACAGCAGCAAGATTACAG 1341
Db 601 AGAAGTACTGCGGTGAGAGGTCCCAAGTTGTGGTGAGCAGCAACAGCAGCAAGATTACAG 660
QY 1342 TCCACTTCAATTTGATCATCTGTHACAAGGACACCGGGTTCTAGCTGAGTACCTTCTCT 1401
Db 661 TCCACTTCAATTTGATCATCTGTHACAAGGACACCGGGTTCTAGCTGAGTACCTTCTCT 720
QY 1402 ACGACTTCCAAACGACCGTCCCGAGGATTTTCATGTGCAAGA 1443
Db 721 ACGACTTCCAAACGACCGTCCCGAGGATTTTCATGTGCAAGA 762
```

RESULT 9
CB208872
LOCUS

DEFINITION AGENCOURT 11350695 NIH MGC 164 Mus musculus cDNA clone
IMAGE:30241011 5', mRNA sequence.
ACCESSION CB208872
VERSION CB208872.2 GI:29133712
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 874)
NIH-MGC <http://mgi.nci.nih.gov/>
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL unpublished
COMMENT On Feb 4, 2003 this sequence version replaced gi:28250435.
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Dr. David Rowe and Dr. Mina
cDNA Library Preparation: Invitrogen Corp
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: NDAM0308 row: g column: 04
High quality sequence stop: 655.
Location/Qualifiers
1. 874
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="IMAGE:30241011"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH MGC 164"
/note="Vector: pCMV-SPORT6.1.cdb; Site 1: EcoRV; Site 2:
NotI; Non-normalized full-length enriched library from
pooled mouse embryonic limb, maxilla and mandible, day
10.5 and 11.5 (size selected for the 0.5-1 kb fragments)
Cloned directionally, priming method: Oligo-dT. cDNA
enrichment: >1k bp, Average insert size 1.8k bp. Priming
sequence: 5'GACTAGTTCTAGATCGAGCGCGCC(T) 3'. Tissue
contributed by, David Rowe. Library constructed by ResGen,
Invitrogen Corp."

BASE COUNT 193 a 230 c 260 g 190 t 1 others

Query Match 24.4%; Score 759.4; DB 14; Length 874;
Best Local Similarity 98.9%; Pred. No. 9.8e-139;
Matches 785; Conservative 0; Mismatches 7; Indels 2; Gaps 2;
QY 1387 CTGAGTACCTCTCTACGACTCCAAACGACCCGTCGCCAGGATGTTTATGTGCAAGCTG 1446
Db 1 CTGAGTACCTCTCTACGACTCCAAACGACCCGTCGCCAGGATGTTTATGTGCAAGCTG 60
QY 1447 GACCGTGATCCGAAGAACTGCGCTCGACGGCTGGCGAGACTGCCCGATTATCTG 1506
Db 61 GACCGTGATCCGAAGAACTGCGCTCGACGGCTGGCGAGACTGCCCGATTATCTG 120
QY 1507 ATGAGCGTTTACTGCGGATGCAATGCCACCCAGCTTCCAGTGCAGAAACACGAGTTCTGCA 1566
Db 121 ATGAGCGTTTACTGCGGATGCAATGCCACCCAGCTTCCAGTGCAGAAACACGAGTTCTGCA 180
QY 1567 AGCCCTCTTCTGGGTCTGTGACAGTGTCAACAGTGTGGGGAAGGAAAGTGCAGGAGG 1626
Db 181 AGCCCTCTTCTGGGTCTGTGACAGTGTCAACAGTGTGGGGAAGGAAAGTGCAGGAGG 240
QY 1627 GCTCAGCTGTCTGCTGGGAGTTTCAAGTCTTCCAAATGGGAAGTGTCTCCTCAGAGCC 1686
Db 241 GCTCAGCTGTCTGCTGGGAGTTTCAAGTCTTCCAAATGGGAAGTGTCTCCTCAGAGCC 300
QY 1687 AGAAGTGTAAATGGGAAGGACAACTGTGGAGATGGTCTGACGAGGCTTTCATGTGACAGCG 1746
Db 301 AGAAGTGTAAATGGGAAGGACAACTGTGGAGATGGTCTGACGAGGCTTTCATGTGACAGCG 360

VERSION
KEYWORDS
SOURCE
ORGANISM

BY756896.1 GI:27190109

EST. Mus musculus (house mouse)

Mus musculus

Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 746)

REFERENCE
AUTHORS

Okazaki, Y., Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S.,
Nikaido, I., Osato, N., Saito, R., Suzuki, H., Yamanaka, I., Kiyosawa, H.

Gojobori, T., Baldarelli, R., Hill, D. P., Bult, C., Hume, D. A.,

Quackenbush, J., Schriml, L. M., Kanapin, A., Matsuda, H., Batalov, S.,

Beisel, K. W., Blake, J. A., Bradt, D., Brusic, V., Chothia, C., Corbani

, L. E., Cousins, S., Dalla, E., Dragani, T. A., Fletcher, C. F., Forrest

, A., Frazer, K. S., Gaasterland, T., Gariboldi, M., Gissi, C., Godzik, A.

, Gough, J., Grimmond, S., Gustincich, S., Hirokawa, N., Jackson, I. J.,

Jarvis, E. D., Kanai, A., Kawaji, H., Kawasawa, Y., Kedzierski, R. M.,

King, B. B., Kohagawa, A., Kurochkin, I. V., Lee, Y., Lenhard, B., Lyons

, P. A., Maglott, D. R., Maltais, L., Marchionni, L., McKenzie, L., Miki

, H., Nagashima, T., Numata, K., Okido, T., Pavan, W. J., Perteau, G.,

Pesole, G., Petrovsky, N., Pallal, R., Pontius, J. U., Qi, D.,

Ramchandran, S., Ravasi, T., Reed, J. C., Reed, D. J., Reid, J., Ring

, B. Z., Ringwald, M., Sandelin, A., Schneider, C., Semple, C. A., Setou

, M., Shimada, K., Sulcane, R., Takenaka, Y., Taylor, M. S., Teasdale

, R. D., Tomita, M., Verardo, R., Wagner, L., Wahlestedt, C., Wang, Y.,

Watanabe, Y., Wells, C., Wilming, L. G., Wynshaw-Boris, A., Yanagisawa

, M., Yang, I., Yang, L., Yuan, Z., Zavolan, M., Zhu, Y., Zimmer, A.,

Carninci, P., Hayatsu, N., Hirozane-Kishikawa, T., Konno, H., Nakamura

, M., Sakazume, N., Sato, K., Shiraki, T., Waki, K., Kawai, J., Aizawa, K.

, Arakawa, T., Fukuda, S., Hara, A., Hashizume, W., Imotani, K., Ishii

, Y., Itoh, M., Kagawa, I., Miyazaki, A., Sakai, K., Sasaki, D., Shibata

, K., Shingawa, A., Yasunishi, A., Yoshino, M., Waterston, R., Lander

, E. S., Rogers, J., Birney, E. and Hayashizaki, Y.

Analysis of the mouse transcriptome based on functional annotation

of 60,770 full-length cDNAs

Nature 420, 563-573 (2002)

22354683

12466851

12466851

Contact: Yoshihide Hayashizaki

Laboratory for Genome Exploration Research Group, RIKEN Genomic

Sciences Center (GSC), Yokohama Institute

The Institute of Physical and Chemical Research (RIKEN)

1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan

Tel: 81-45-503-9222

Fax: 81-45-503-9216

Email: genome-res@gs.riken.go.jp,

URL: http://genome.gsc.riken.go.jp/

Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Carninci, P., Fukuda

, S., Hashizume, W., Hayashida, K., Hirozane, T., Hori, F., Imotani, K.,

Ishii, Y., Itoh, M., Kagawa, I., Kawai, J., Kojima, Y., Kondo, S., Konno

, H., Koya, S., Miyazaki, A., Murata, M., Nakamura, M., Nomura, K.,

Numazaki, R., Ohno, M., Osato, N., Saito, R., Sakazume, N., Sano, H.,

Sasaki, D., Sato, K., Shibata, K., Shiraki, T., Tagami, M., Takeda, Y.,

Waki, K., Watahiki, A., Muramatsu, M. and Hayashizaki, Y. Direct

Submission

Computational Analysis of Full-Length Mouse cDNAs Compared with

Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)

Normalization and subtraction of cap-trapper-selected cDNAs to

prepare full-length cDNA libraries for rapid discovery of new

genes. Genome Res. 10 (10), 1617-1630 (2000)

RIKEN integrated sequence analysis (RISA) system--384-format

sequencing pipeline with 384 multicapillary sequencer. Genome Res.

10 (11), 1757-1771 (2000)

Computer-based methods for the mouse full-length cDNA

encyclopedia: real-time sequence clustering for construction of a

nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)

cDNA library was prepared and sequenced in Mouse Genome

Encyclopedia Project of Genome Exploration Research Group in Riken

Genomic Sciences Center and Genome Science Laboratory in RIKEN.

Division of Experimental Animal Research in Riken contributed to

prepare mouse tissues.

Please visit our web site (<http://genome.gsc.riken.go.jp>) for

further details.

FEATURES
source

Location/Qualifiers
1..746

/organism="Mus musculus"

/mol_type="mRNA"

/strain="CS7BL/6J"

/db_xref="taxon:10090"

/clone="I1C0035F06"

/dev_stage="blastocyst"

/clone_lib="RIKEN full-length enriched, blastocyst"

172 a 193 c 224 g 155 t 2 others

BASE COUNT

ORIGIN

Query Match

Best Local Similarity

Matches 741; Conservative

23.8%; Score 739.2; DB 14; Length 746;

99.3%; Pred. No. 9.3e-135;

0; Mismatches 5; Indels

0; Gaps

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

RESULT 12

BI416218

LOCUS

DEFINITION

BI416218 846 bp mRNA linear EST 14-AUG-2001
602989878F1 NCI_CGAP_Lu33 Mus musculus cDNA clone IMAGE:5145762 5',

3088

746

3062

720

3063

3088

3088

3088

3088

3088

3088

3088

3088

3088

3088

```

mRNA sequence.
B1416218
BI416218.1 GI:15177141
EST.
Mus musculus
Mus musculus (house mouse)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
NTH-MGC http://mgi.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Gilbert Smith, Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
Bonaldo, Ph.D.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM1359 Row: 1 column: 19
High quality sequence start: 20
High quality sequence stop: 793.
Location/Qualifiers
1. .846
/organism="Mus musculus"
/mol_type="mRNA"
/strain="CZECH II"
/db_xref="taxon:10090"
/clone="IMAGE:5145762"
/tissue_type="pooled lung tumors"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NCI CGAP Lu33"
/note="Organ: Lung; Vector: p773D-Pac (Pharmacia) with a
modified polylinker; Site: NotI; Site 2: EcoRI; lsc
strand cDNA was prepared from mRNA obtained from pooled
lung tumors with a Not I - oligo(dT) primer [5',
TGTTACCAATCTGAAGTGGGAGCGCGCTCTCTTTTCTTTTCTTTT 3']
Double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not
I and Eco RI sites of the modified p773 vector. Library
went through one round of normalization, and was
constructed by Bento Soares and M. Fatima Bonaldo. "
BASE COUNT 178 a 224 c 262 g 182 t
ORIGIN
Query Match 23.8%; Score 739.2; DB 12; Length 846;
Best Local Similarity 97.1%; Pred. No. 9e-135;
Matches 816; Conservative 0; Mismatches 18; Indels 6; Gaps 6;
QY 1687 AGAAGTGTAAATGGGAAGGACAACTGTGAGATGGGTCTGACGAGGCTTCATGTGACAGCG 1746
DB 1 AGATCCGACAGATGAGGACACTGTGAGATGGGTCTGACGAGGCTTCATGTGACAGCG 60
QY 1747 TGAATGTCGTCTTTGCCAATAATACCTACCGCTGCCAAATGGCCCTCTGTCTGAGCA 1806
DB 61 TGAATGTCGTCTTTGCCAATAATACCTACCGCTGCCAAATGGGCTCTGTCTGAGCA 120
QY 1807 AGGGCAACCTGAGTGTGATGGGAAGCAGCACTGTAGCGATGGCTCCGATGAGAAAAC 1866
DB 121 AGGGCAACCTGAGTGTGATGGGAAGCAGCACTGTAGCGATGGCTCCGATGAGAAAAC 180
QY 1867 GTGACTGTGGGCTGGATCTTTACAAACAGGCTCGGTGGTGGTGGCAGCAATGCGG 1926
DB 181 GTGACTGTGGGCTGGATCTTTACAAACAGGCTCGGTGGTGGTGGCAGCAATGCGG 240
QY 1927 ACGAGGCGAGTGGCCCTGGCAGGTGAGCTTCCACGCCCTGGGCGAGGCCCACTTGTGTG 1986
DB 241 ACGAGGCGAGTGGCCCTGGCAGGTGAGCTTCCACGCCCTGGGCGCA-GGCCACTTGTGTG 299
QY 1987 GGGCCCTCGCTACTCTCTCTGACTGGTGGTCTCTGACAGCTCATGTCTTCAGGATGACA 2046

```

Db 300 GGGCCCTCGCTCATCTCTCTCTGAGCTGGTCTCTGAGCTCATTTTTCAGGATGACA 359
 QY 2047 AATATTTCAAGTACTCAGACTACAGTGTGAGCGCTTCTCTGGGTCTCTGACGAGCA 2106
 Db 360 AATATTTCAAGTACTCAGACTACAGTGTGAGCGCTTCTCTGGGTCTCTGACGAGCA 419
 QY 2107 GCAAGCGCAGTGGCTCTGGGGTGCAGAGCTGAAGCTCAAACTATCATCACCCACCTT 2166
 Db 420 GCAAGCGCAGTGGCTCTGGGGTGCAGAGCTGAAGCTCAAACTATCATCACCCACCTT 479
 QY 2167 CTTCAATGATTTTCACTTTCGACTATGACATCGCTTGTCTGGAGCTGGAGAGTCGCTG 2226
 Db 480 CTTCAATGATTTTCACTTTCGACTATGACATCGCTTGTCTGGAGCTGGAGAGTCGCTG 539
 QY 2227 AGTACAGCACCGTCTGTCGGGCCCATCTGCTGCTGATGCTACCCATGTCTTCCCTCTG 2286
 Db 540 AGTACAGCACCGTCTGTCGGGCCCATCTGCTGCTGATGCTACCCATGTCTTCCCTCTG 599
 QY 2287 GC-AAGGCCATCTGGGTTCACAGCTGGGGGCACACAAAAGAGGAGGT-ACCGAGCGCT 2344
 Db 600 GCAAGGCCATCTGGGTTCACAGCTGGGGGCACACAAAAGAGGAGGTTCACCGAGCGCT 659
 QY 2345 GATCTTCACAGAGGTCAGATCCGTGTCTATC-RACAGAGCACTGTGAGAGCTCATGC 2403
 Db 660 GATCTTCACAGAGGTCAGATCCGTGTCTATCAAAACAGAGCGACTGTGAGAGCTCATGC 719
 QY 2404 CGCAGCAGATCACCCACAGATGATGTGTGGGTTCCTCAGTGGGGTGTGGACTCCT 2463
 Db 720 CGCAG-AGATCACCCACAGATGATGTGTGGGTTCCTCAGTGGGGTGTGGACTCCT 778
 QY 2464 GCGAGGTGATCTCTGGTGGCCCTGTCAAGCGCGGAGAGATGGCGAATGTTCAGG 2523
 Db 779 G-CAGGGTGAATCTCTGGTGGCCCTGTCAAGCGCGGAGAGATGGCGAATGTTCAGG 837

RESULT 13
 BF301347 1116 bp mRNA linear EST 21-NOV-2000
 LOCUS 602029796F1 NCI_CGAP_SG2 Mus musculus cDNA clone IMAGE:4165103 5',
 DEFINITION mRNA sequence.
 ACCESSION BF301347
 KEYWORDS EST.
 SOURCE Mus musculus
 ORGANISM Mus musculus (house mouse)
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 NIH-MGC http://mgi.nci.nih.gov/
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished
 Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-r@mail.nih.gov
 Tissue Procurement: Jeffrey E. Green, M.D.
 cDNA Library Preparation: Life Technologies, Inc.
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: LLAM9451 Row: 0 column: 24
 High quality sequence stop: 686.
 Location/Qualifiers
 1. .1116
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="FVB/N"
 /db_xref="taxon:10090"
 /clone="IMAGE:4165103"
 /lab_host="DH10B (T1 phage-resistant)"
 /clone_lib="NCI CGAP SG2"
 /note="Organ: salivary gland; Vector: pCMV-SPORT6; Site_1:

FEATURES
 source
 1. .1116
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="FVB/N"
 /db_xref="taxon:10090"
 /clone="IMAGE:4165103"
 /lab_host="DH10B (T1 phage-resistant)"
 /clone_lib="NCI CGAP SG2"
 /note="Organ: salivary gland; Vector: pCMV-SPORT6; Site_1:

Not1; Site 2: SalI; Cloned unidirectionally. Primer: Oligo
dT. Average insert size 1.3 kb. Constructed by Life
Technologies. Note: this is a NCI_CGAP Library."

```
BASE COUNT      325 a      315 c      302 g      174 t
ORIGIN
Query Match      23.7%; Score 735.8; DB 10; Length 1116;
Best Local Similarity 92.1%; Pred. No. 3.9e-134;
Matches 800; Conservative 0; Mismatches 62; Indels 7; Gaps 2;
QY 1137 CACTACCCGCCCAACATCAATGTCACATGGAATCAAGTGGCCCAACCAACCGGAACGTG 1196
DB 1 CACTACCCGCCCAACATCAATGTCACATGGAATCAAGTGGCCCAACCAACCGGAACGTG 60
QY 1197 AAGGTGGCCTTCAACTCTTCTATCTGTGGACCCCAACGACGAGTGGCTCTCGACCC 1256
DB 61 AAGGTGGCCTTCAACTCTTCTATCTGTGGACCCCAACGACGAGTGGCTCTCGACCC 120
QY 1257 AAGGACTATGTGGAGATCAACGGGGAGAGTACTCGGTGAGAGGTCCAGTTTGTGGTG 1316
DB 121 AAGGACTATGTGGAGATCAACGGGGAGAGTACTCGGTGAGAGGTCCAGTTTGTGGTG 180
QY 1317 AGCAGCAACAGCAGCAAGATTACGTCCACTTCCATTCTGATCACTGTACAGGACAC 1376
DB 181 AGCAGCAACAGCAGCAAGATTACGTCCACTTCCATTCTGATCACTGTACAGGACAC 240
QY 1377 GGGTTCCTAGCTGAGTACCTCTCTACGACTCCACGACCCGTCGCCAGGGATGTCATG 1436
DB 241 GGGTTCCTAGCTGAGTACCTCTCTACGACTCCACGACCCGTCGCCAGGGATGTCATG 300
QY 1437 TGCAAGACTGAGCGGTGATCCGAAAGAACTGCGTGCAGCGGTGGCGAGACTGCCCG 1496
DB 301 TGCAAGACTGAGCGGTGATCCGAAAGAACTGCGTGCAGCGGTGGCGAGACTGCCCG 360
QY 1497 GATTATAGTATGAGCGTTACTGCCGATGCCAATGCCACCCACGATTCACGTGCAAAAC 1556
DB 361 GATTATAGTATGAGCGTTACTGCCGATGCCAATGCCACCCACGATTCACGTGCAAAAC 420
QY 1557 CAGTTCTGCAAGCCCTCTTCTGGGTCTGTGACAGTGTCAAGCACTGTGGGACCGAAGT 1616
DB 421 CAGTTCTGCAAGCCCTCTTCTGGGTCTGTGACAGTGTCAAGCACTGTGGGACCGAAGT 480
QY 1617 GACGAGGAGGCTGACGCTGCTCTGCGAGTTTCAAGTGTTCGAATGGGAAGTGTCTC 1676
DB 481 GACGAGGAGGCTGACGCTGCTCTGCGAGTTTCAAGTGTTCGAATGGGAAGTGTCTC 540
QY 1677 CCTCAGAGCCAGAGTGTAAATGGGAAGGACACTGTGGAGATGGGTCTGACAGGCTTCA 1736
DB 541 CCTCAGAGCCAGAGTGTAAATGGGAAGGACACTGTGGAGATGGGTCTGACAGGCTTCA 600
QY 1737 TGTGACAGCGTGAATGTCTCTCTTGACCAATATACCTACCGCTGCCAAATGGCCTC 1796
DB 601 TGTGACAGCGTGAATGTCTCTCTTGACCAATATACCTACCGATGCCAAATGGCCTC 660
QY 1797 TGTCTGAGCAAGGGCAACCCCTGAGTGTGTATGGGAAGCAAGCACTGTAGCGATGGCTCCGAT 1856
DB 661 TGTCTGAGCAAGGGCAACCCCTGAGTGTGTATGGGAAGCAAGCACTGTAGCGATGGCTCCGCA 720
QY 1857 G---AGAAAACTGTGACTGTGGCTGCGATCTTTTACCAACAGGCTGCGGTGTGTG 1913
DB 721 GGACAAACACACTGGGACTGTGGGACGCGATCTTTTACCAACAGGATCAGGTGACAGGT 780
QY 1914 GGCACGAATGCG---GACGAGGGCGAGTGGCCCTGGCAGGTGAGCTCCAGCCCTGGG 1969
DB 781 GGAACGAATTCAGACAAGGAACAAAGTGGCCGTGCACGAGGCTCCAGGCTGTG 840
QY 1970 CCAGGCGCACTGTGTGGGCGCTCGCTCA 1998
DB 841 GAACAGCGCAATGAGTGGGAACCCCTAGCA 869
```

RESULT 14
BI646380

LOCUS BI646380 786 bp mRNA linear EST 12-SEP-2001
DEFINITION 603276244F1 NCI_CGAP_Mam3 Mus musculus cDNA clone IMAGE:5316539 5',
mRNA sequence.
ACCESSION BI646380
VERSION BI646380.1 GI:15560616
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 786)
NIH-MGC <http://mgc.nci.nih.gov/>.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Lohar Hennighausen Ph.D., Chu-Xia Deng Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: LHAM11802 row: h column: 12
High quality sequence stop: 764.

FEATURES

Location/Qualifiers
1..786
/organism="Mus musculus"
/mol_type="mRNA"
/strain="129,C57BL/6J,FVB/N"
/db_xref="taxon:10090"
/clones="IMAGE:5316539"
/tissue_type="tumor, gross tissue"
/dev_stage="10 months"
/lab_host="DH10B"
/clone_libs="NCI_CGAP Mam3"
/note="Organ: mammary; Vector: pCMV-SPORT6; Site 1: SalI;
Site 2: NotI; Cloned unidirectionally. Primer: Oligo dT.
Library constructed by Life Technologies. Investigators
providing samples: Lohar Hennighausen/Chu-Xia Deng, NIH
Reference for transgenic model: Xu et al., Nature Genetics
22, 37-43 (1999)."
22..37-43 (1999)."

BASE COUNT 177 a 203 c 235 g 171 t
ORIGIN

Query Match 23.6%; Score 732; DB 12; Length 786;
Best Local Similarity 98.7%; Pred. No. 2.4e-133;
Matches 780; Conservative 0; Mismatches 5; Indels 5; Gaps 4;
QY 1409 CAACGACCCGTCGCCCGGATGTTTCATGTCAAGACTGACGGTGCATCCGAAAGAACT 1468
DB 1 CAACGACCCGTCGCCCA-GGATGTTTCATGTCAAGACTGACGGTGCATCCGAAAGAACT 59
QY 1469 GGGCTGCGACCGCTGGGCGAGCTGCCGGATTATAGTATGATGAGCGTTACTGCCGATCAA 1528
DB 60 GGGCTGCGACCGCTGGGCGAGCTGCCGGATTATAGTATGATGAGCGTTACTGCCGATCAA 119
QY 1529 TGCCACCCACAGTTTCAGCTGCAAAAACAGTTTCGCAAGCCCTCTTCTGGGTCTGTGA 1588
DB 120 TGCCACCCACAGTTTCAGCTGCAAAAACAGTTTCGCAAGCCCTCTTCTGGGTCTGTGA 179
QY 1589 CAGTGTCAACGACTGTGGGACCGGAAGTGACGAGGCGGTGCAGCTGTCTCTGTGGGAG 1648
DB 180 CAGTGTCAACGACTGTGGGACCGGAAGTGACGAGGCGGTGCAGCTGTCTCTGTGGGAG 239
QY 1649 TTTCAGTGTTCGAATGGGAAGTGTCTCCCTCAGAGCCAGAGTGTAAATGGGAAGCAAA 1708
DB 240 TTTCAGTGTTCGAATGGGAAGTGTCTCCCTCAGAGCCAGAGTGTAAATGGGAAGCAAA 299
QY 1709 CTGTGGAGATGGGTCTGACGAGGCTTCATGTGACAGCGTGAATGTCTCTTTGACCAAA 1768
DB 300 CTGTGGAGATGGGTCTGACGAGGCTTCATGTGACAGCGTGAATGTCTCTTTGACCAAA 359

```

1769 ATATACCTACCGCTGCACAAATGCGCTCTGTCTGAGCAAGGCAACCCCTGAGTGTGATGG 1828
Db      |||
360 ATATACCTACCGCTGCACAAATGCGCTCTGTCTGAGCAAGGCAACCCCTGAGTGTGATGG 419
Qy      |||
1829 GAAGACGAGCTGTAGCGATGCGCTCCGATGAGAAAACTGTGACTGTGGCTGCGATCCTT 1888
Db      |||
420 GAAGACGAGCTGTAGCGATGCGCTCCGATGAGAAAACTGTGACTGTGGCTGCGATCCTT 479
Qy      |||
1889 TACCAACAGCGCTCGCTGGTGTGGTGACAGAAATGCGGACGAGCGAGTGGCCCTCGCA 1948
Db      |||
480 TACCAACAGCGCTCGCTGGTGTGGTGACAGAAATGCGGACGAGCGAGTGGCCCTCGCA 539
Qy      |||
1949 GGTGAGCCTCCAGCCCTCGGCGCCAGGCGCACTTGTGTGGGCGCTCGCTCATCTCTCTGA 2008
Db      |||
540 GGTGAGCCTCCAGCCCTCGGCGCCAGGCGCACTTGTGTGGGCGCTCGCTCATCTCTCTGA 599
Qy      |||
2009 CTGCTGTCTCTGAGCTCATCTGCTTTCAGGATGACAAAAATTTCAAGTACTCAGACTA 2068
Db      |||
600 CTGCTGTCTCTGAGCTCATCTGCTTTCAGGATGACAAAAATTTCAAGTACTCAGACTA 659
Qy      |||
2069 CAAG-ATGTGACGCGCTTCTGCTGGTCTGCTGGACGAGCAAGCGAGTGCCTCTCGGG 2127
Db      |||
660 CACGAATGTGACGCGCTTCTGCTGGTCTGCTGGACGAGCAAGCGAGTGCCTCTCGGG 719
Qy      |||
2128 TCGAGGAGCTGAAGCTCAAACTATCATCATCACCACCTTCTTCAATGATTTCACTTCG 2187
Db      |||
720 TCGAGGAGCTGAAGCTCAAACTATCATCATCACCACCTTCTTCAATGATTTCACTTCG 776
Qy      |||
2188 ACTATGACAT 2197
Db      |||
777 ACTATGACAT 786

RESULT 15
BG915045
LOCUS      602814325F1 NCI_CGAP_Mam4 Mus musculus cDNA clone IMAGE:4936590 5',
DEFINITION mRNA sequence.
ACCESSION BG915045
VERSION   BG915045.1 GI:14295521
KEYWORDS  EST.
SOURCE    Mus musculus (house mouse)
ORGANISM  Mus musculus
REFERENCE 1 (bases 1 to 928)
AUTHORS   NIH-MGC http://mgc.nci.nih.gov/
TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL   Unpublished
COMMENT   Contact: Robert Strausberg, Ph.D.
          Email: cgabbs-r@mail.nih.gov
          Tissue Procurement: Lothar Hennighausen Ph.D., Priscilla Furth
          Ph.D.
          cDNA Library Preparation: Life Technologies, Inc.
          cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
          DNA Sequencing by: Incyte Genomics, Inc.
          Clone distribution: MGC clone distribution information can be
          found through the I.M.A.G.E. Consortium/LLNL at:
          http://image.llnl.gov
          Plate: LLAM10869 row: a column: 07
          High quality sequence stop: 793.
          Location/Qualifiers
            1..928
              /organism="Mus musculus"
              /mol_type="mRNA"
              /strain="NMRI"
              /db_xref="taxon:10090"
              /clone="IMAGE:4936590"
              /issue_type="tumor, gross tissue"
              /dev_stage="5 months"
              /lab_host="DH10B"
              /clone_lib="NCI_CGAP_Mam4"
              /note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: SalI;

```

```

Site 2: NotI; Cloned unidirectionally. Primer: Oligo dr.
Library constructed by Life Technologies. Investigators
providing samples: Lothar Hennighausen/Priscilla Furth,
NIH Reference for transgenic model: Li et al., Cell Growth
and Differentiation 7, 3-11 (1996).
BASE COUNT 216 a 258 c 273 g 181 t
ORIGIN
Query Match 23.2%; Score 721; DB 12; Length 928;
Best Local Similarity 93.2%; Pred. No. 3.3e-131;
Matches 832; Conservative 0; Mismatches 50; Indels 11; Gaps 7;
1 CTTTACCAACAGGCTCGCTGGTGTGGTGCGACGAATGCGGACGAGGCGAGTGGCCCTG 1945
Db      |||
1886 CTTTACCAACAGGCTCGCTGGTGTGGTGCGACGAATGCGGACGAGGCGAGTGGCCCTG 60
Qy      |||
1946 GCAGGTGAGGCTCCAGCCCTCGGCGCCAGGCGCACTTGTGTGGGCGCTCGCTCATCTCTCC 2005
Db      |||
61 GCAGGTGAGGCTCCAGCCCTCGGCGCA-GGCCCACTTGTGTGGGCGCTCGCTCATCTCTCC 119
Qy      |||
2006 TGACTGGCTGCTCTCTGAGCTCATCTGCTTTCAGGATGACAAAAATTTCAAGTACTCAGA 2065
Db      |||
120 TGACTGGCTGCTCTCTGAGCTCATCTGCTTTCAGGATGACAAAAATTTCAAGTACTCAGA 179
Qy      |||
2066 CTACACGATGTGAGCGGCTTCTGCTGGTCTGCTGGACGAGCAAGCGAGTGCCTCTGG 2125
Db      |||
180 CTACACGATGTGAGCGGCTTCTGCTGGTCTGCTGGACGAGCAAGCGAGTGCCTCTGG 239
Qy      |||
2126 GGTGAGGAGCTGAGCTCAAACTATCATCATCACCACCTTCTTCAATGATTTCACTT 2185
Db      |||
240 GGTGAGGAGCTGAGCTCAAACTATCATCATCACCACCTTCTTCAATGATTTCACTT 299
Qy      |||
2186 CGACTATGACATCGCTTGTGTGAGCTGAGAACTGCTGGTGGAGTACAGCACCGCTGTCG 2245
Db      |||
300 CGACTATGACATCGCTTGTGTGAGCTGAGAACTGCTGGTGGAGTACAGCACCGCTGTCG 359
Qy      |||
2246 CCCCATCTGCTGCTGATGCTACCAATGCTTCCCTGCTGGCAAGGCATCTGGTTCAC 2305
Db      |||
360 CCCCATCTGCTGCTGATGCTACCAATGCTTCCCTGCTGGCAAGGCATCTGGTTCAC 419
Qy      |||
2306 AGGCTGGGGGCACAAAAAGAGGAGGTACCGGAGCGCTGATCTCGAAGAGGCTGAGAT 2365
Db      |||
420 AGGCTGGGGGCACAAAAAGAGGAGGTACCGGAGCGCTGATCTCGAAGAGGCTGAGAT 479
Qy      |||
2366 CCGTGTCAATCAACAGACCACTCTGTGAGGAGCTCATCGCGAGAGTACACCCACGAAT 2425
Db      |||
480 CCGTGTCAATCAACAGACCACTCTGTGAGGAGCTCATCGCGAGAGTACACCCACGAAT 539
Qy      |||
2426 GATGTGTGGGTTTCTCTAGTGGGGGTGTGGACTCTGCTGCGAGGCTGACTCTGGT 2483
Db      |||
540 GATGTGTGGGTTTCTCTAGTGGGGGTGTGGACTCTGCTGCGAGGCTGACTCTGGTGGCC 599
Qy      |||
2484 CCCTTGTCAAGCGCGGAGAAAGATGGGCGAATGTTCCAGGCTGGTGTGGTGGAGTGGGGT 2543
Db      |||
600 CTTGTCAAGCGCGGAGAAAGATGGGCGAATGTTCCAGGCTGGTGTGGTGGAGTGGGGT 659
Qy      |||
2544 GAAGGCTGCGCTCAGAGGAAACAG-CCAGGCGGTGTACA-CAAGGCTCCTGTAGTTGGG 2601
Db      |||
660 GAAGGCTGCGCTCAGAGGAAACAGCCAGGCGGTGTACACCAAGGCTCCTGTAGTTGGG 719
Qy      |||
2602 ACTGGATCAAAAGACACACTGGGGTATAGCAGATGGA-CAGACGCGGACCAACAAACAC 2660
Db      |||
720 AACTGGATCAAAAGACACACTGGGGTATAGCAGATGGA-CAGACGCGGACCAACAAACAC 779
Qy      |||
2661 CCACAGGAGTCCCGACATGCACACCTGGAT---ACAGGAGGAGAACTGACGACATTT 2717
Db      |||
780 CCACAGGAGTCCCGAATGACACCTGGATTTACAGGAGCAGGAACTGACGACATCA 839
Qy      |||
2718 ATGCTGTGGCTCCCCCCCCCAACACACACCGAGACTGTGAATCTGATCCTTAG 2770
Db      |||
840 TGCTTGTGGCATCCACCCCC---AAAAACAACAACACTGTGAACCTGCTACCTAAG 890

```

Search completed: February 21, 2004, 07:12:09
Job time : 6230 secs